DOCUMENT RESUME

ED 424 982 PS 027 162

AUTHOR	Zaslow, Martha J., Ed.; Eldred, Carolyn A., Ed.
TITLE	Parenting Behavior in a Sample of Young Mothers in Poverty: Results of the New Chance Observational Study.
INSTITUTION	Manpower Demonstration Research Corp., New York, NY.
SPONS AGENCY	Foundation for Child Development, New York, NY.; Grant (W.T.) Foundation, New York, NY.; Smith Richardson Foundation, Inc., Greensboro, NC.; National Inst. of Child Health and Human Development (NIH), Bethesda, MD.
PUB DATE	1998-04-00
NOTE	465p.
PUB TYPE	Collected Works - General (020) -- Reports - Research (143)
EDRS PRICE	MF01/PC19 Plus Postage.
DESCRIPTORS	Attachment Behavior; Child Development; *Child Rearing; Demonstration Programs; Intergenerational Programs; Job Skills; Job Training; Longitudinal Studies; Low Income Groups; *Mothers; *Parent Child Relationship; Participant Observation; Poverty; Program Evaluation; Tables (Data); Welfare Recipients
IDENTIFIERS	Cross Generational Studies; *New Chance; Relationship Quality

ABSTRACT

This study examines parenting in 290 of the 2,322 families studied in the New Chance Demonstration, a national research and demonstration program operated between 1989 and 1992 in 10 states. The demonstration project sought to improve the economic prospects and overall well-being of low-income young mothers (ages 16 to 22) and their children through comprehensive, intensive services and to further the development and promote the health of the participants' children. The 14 chapters in this monograph are: (1) "Introduction: The Context for the New Chance Observational Study" (Zaslow); (2) "The Methodology of the New Chance Observational Study" (Morrison, Eldred, Zaslow, and Dion); (3) "Participation in Program Components That Could Affect Parenting Behavior" (Dion, Zaslow, and Morrison); (4) "The Affective Quality of Mother-Child Interaction" (Weinfield, Egeland, and Ogawa); (5) "Mother-Child Interactions Related to the Emergence of Literacy" (De Temple and Snow); (6) "Completing the Portrayal of Parenting Behavior with Interview-Based Measures" (Morrison, Zaslow, and Dion); (7) "Integration: Looking across the Differing Measures of Parenting" (Zaslow, Dion, and Morrison); (8) "Parenting in a Broader Context: An Examination of the Multiple Influences on Child Outcomes" (New Chance Observational Study Research Team); (9) "Key Findings and Their Implications" (Zaslow); (10) "Expanding the Methodological Horizons of Child Development and Survey Research" (Eldred); (11) "Implementing Observational Research Within a Survey Context" (Eldred); (12) "Findings on the Administration of the Observational Session" (Eldred); (13) "An Assessment of the Data Collection Effort and Lessons for Future Research Efforts" (Eldred); and (14) "Measurement Implications of the New Chance Observational Study" (Eldred). (Contains 87 tables and 176 references.) (KB)

ED 424 982

Parenting Behavior in a Sample of Young Mothers in Poverty

Results of the New Chance Observational Study

Martha J. Zaslow
Carolyn A. Eldred
Editors

MDRC

April 1998

Manpower Demonstration
Research Corporation

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY
Carol A. Emig
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
1

PS 027162

BEST COPY AVAILABLE

BOARD OF DIRECTORS

ROBERT REISCHAUER, *Chairman*
Senior Fellow
Brookings Institution

PAUL H. O'NEILL, *Treasurer*
Chairman and CEO
Alcoa

MARY JO BANE
Professor of Public Policy
John F. Kennedy School of Government
Harvard University

ANTONIA HERNANDEZ
President and General Counsel
Mexican American Legal Defense and
Educational Fund

JAMES H. JOHNSON, JR.
E. Maynard Adams Professor of Business,
Geography, and Sociology
Director, Urban Investment Strategies Center
University of North Carolina

ANNA KONDRATAS
Senior Associate
Urban Institute

RICHARD J. MURNANE
Professor of Education
Graduate School of Education
Harvard University

RUDOLPH G. PENNER
Senior Fellow
Urban Institute

MARION O. SANDLER
Chairman and CEO
Golden West Financial Corporation and
World Savings and Loan Association

ISABEL V. SAWHILL
Senior Fellow
Brookings Institution

ROBERT SOLOW
Institute Professor
Massachusetts Institute of Technology

MITCHELL SVIRIDOFF
Professor Emeritus and Senior Fellow
Community Development Research Center
New School for Social Research

WILLIAM JULIUS WILSON
Malcolm Wiener Professor of Social Policy
John F. Kennedy School of Government
Harvard University

JUDITH M. GUERON
President
Manpower Demonstration Research Corporation

MDRC

Parenting Behavior in a Sample of Young Mothers in Poverty

Results of the New Chance Observational Study

Martha J. Zaslow
Carolyn A. Eldred
Editors

MDRC

April 1998

Manpower Demonstration
Research Corporation

Funders of the New Chance Observational Study

Foundation for Child Development	Smith Richardson Foundation
William T. Grant Foundation	National Institute of Child Health and Human Development

Dissemination of MDRC publications is also supported by MDRC's Public Policy Outreach funders: the Ford Foundation, the Ambrose Monell Foundation, the Alcoa Foundation, and the James Irvine Foundation.

The findings and conclusions presented in this report do not necessarily represent the official positions or policies of the funders.

Copyright © 1998 by the Manpower Demonstration Research Corporation

Contents

Foreword xi
Acknowledgments xiv
Executive Summary ES-1
by Martha J. Zaslow and Carolyn A. Eldred

Part I: The New Chance Observational Study

1 Introduction: The Context for the New Chance Observational Study
by Martha J. Zaslow 1

I. Goals of the New Chance Observational Study 1
II. The New Chance Intervention 5
III The New Chance Evaluation 9
IV. Questions Addressed by the New Chance Observational Study 20
V. The Organization of the Monograph 23

2 The Methodology of the New Chance Observational Study
by Donna R. Morrison, Carolyn A. Eldred, Martha J. Zaslow, and M. Robin Dion 25

I. The Sample 25
II. Study Procedures 47
III. Strategies for Analyses in This Monograph 50

3 Participation in Program Components That Could Affect Parenting Behavior
by M. Robin Dion, Martha J. Zaslow, and Donna R. Morrison 54

I. Introduction 55
II. Four Perspectives Related to Participation and Parenting Outcomes 56
III. Frequency and Rates of Participation 57
IV. Factors Associated with Extent of Participation by Experimental Group Mothers 64
V. The Nature and Content of the New Chance Parenting Education Component 65
VI. Summary and Conclusions 69

4 The Affective Quality of Mother-Child Interaction
by Nancy S. Weinfield, Byron Egeland, and John R. Ogawa 71

I. The Importance of Studying the Affective Quality of Mother-Child Interaction 71
II. Procedures 76
III. Findings and Discussion 80
IV. Conclusions 112

5 **Mother-Child Interactions Related to the Emergence of Literacy**
by Jeanne De Temple and Catherine Snow 114

I. The Importance of Mother's Talk with Her Child in Language Acquisition and in the Development of Discourse Skills Required in School 115
II. The Coding of Literacy-Related Interactions in the New Chance Observational Study 119
III. Findings 127
IV. Conclusions 168

6 **Completing the Portrayal of Parenting Behavior with Interview-Based Measures**
by Donna R. Morrison, Martha J. Zaslow, and M. Robin Dion 170

I. A Description of the Interview-Based Measures of Parenting 172
II. Variation in the Interview-Based Measures of Parenting According to Baseline and 18-Month Characteristics 176
III. Impacts of New Chance on the Interview-Based Measures of Parenting 200
IV. Summary 211

7 **Integration: Looking Across the Differing Measures of Parenting**
by Martha J. Zaslow, M. Robin Dion, and Donna R. Morrison 213

I. Consistency and Magnitude of Program Impacts on Parenting and Linkages with Program Participation 215
II. The Contribution of Observational Measures of Parenting to the New Chance Evaluation 223

8 **Parenting in a Broader Context: An Examination of the Multiple Influences on Child Outcomes**
by the New Chance Observational Study Research Team 248

I. Analysis Strategies 250
II. Predicting Child Outcomes from the Broader Set of Variables 261
III. Conclusion 268

9 **Key Findings and Their Implications**
by Martha J. Zaslow 269

I. Descriptive Findings on Parenting Behavior: Laying the Groundwork for Examination of Impacts 269
II. Impacts of the New Chance Program on Parenting Behavior 274
III. Different Measures of Parenting Behaviors as Predictors of Child Outcomes 279
IV. The Role of Parenting and Other Important Factors in Shaping Child Outcomes 280

Part II: Methodological Assessment of the New Chance Observational Study

10 Expanding the Methodological Horizons of Child Development and Survey Research
by Carolyn A. Eldred 281

I. The Rationale for Examining the Study from a Methodological Perspective 281
II. The Survey Model as an Option for Research on Parenting and Child Outcomes 284
III. An Overview of the Methodological Issues in Part II 293

11 Implementing Observational Research Within a Survey Context
by Carolyn A. Eldred 297

I. Instrument and Procedures Design 297
II. Data Collection 304

12 Findings on the Administration of the Observational Session
by Carolyn A. Eldred 309

I. The Environment for Conducting the Session 310
II. Interviewer Performance 316
III. Subjects' Experience of the Session 326
IV. Determinants of the Session Experience 336

13 An Assessment of the Data Collection Effort and Lessons for Future Research Efforts
by Carolyn A. Eldred 342

I. Aspects of the Approach That Served the Study Well 342
II. Recommended Refinements to the Approach 344
III. The Feasibility of Observational Research Within a Survey Framework 348

14 Measurement Implications of the New Chance Observational Study
by Carolyn A. Eldred 349

I. Overall Study Design 350
II. The Design of Measures, Instruments, and Procedures 352
III. A Bridge to Future Work 353

Appendices

Appendix A 354
Appendix B 355
Appendix C 356

References 366
Selected Publications on MDRC Projects 376

List of Tables

Table		Page
ES-1	Parenting Measures Included in the New Chance Observational Study	ES-8
1.1	Program Guidelines for Participation and Mean Hours of Actual Participation in New Chance Within 18-Month Follow-Up Period	13
2.1	Selected Characteristics of the New Chance Observational Study Sample at Baseline	30
2.2	Impacts on Educational Attainment, Literacy, and Classes in Education, Training, or Skills at 18-Month and 42-Month Follow-Ups	33
2.3	Impacts on Time Spent Working and AFDC Receipt at 18-Month and 42-Month Follow-Ups	34
2.4	Impacts on Fertility-Related Outcomes at 18-Month and 42-Month Follow-Ups	35
2.5	Impacts on Marital Status and Residence at 18-Month and 42-Month Follow-Ups	36
2.6	Impacts on Child Care Use for the Focal Child at or Within 18-Month and 42- Month Follow-Ups	37
2.7	Impacts on Psychological Characteristics of Respondent at 18-Month and 42-Month Follow-Ups	38
2.8	Impacts on Child Development Outcomes at 42-Month Follow-Up	39
2.9	Impacts on Child Development Outcomes at 42-Month Follow-Up, Teacher Survey Subsample	40
2.10	Impacts on Observational Interview Variables at 21-Month Follow-Up	45
2.11	Sources of Parenting Measures Included in the New Chance Observational Study	51
3.1	Self-Reported Rates of Participation by Experimentals and Controls at 18-Month Follow-Up	58
3.2	Participation of New Chance Observational Study Experimentals in Parenting Classes Within 18-Month Follow-Up Period	60
3.3	Participation of New Chance Observational Study Experimentals in All MIS-Recorded Activities Within 18-Month Follow-Up Period	62
4.1	Teaching Task Rating Scales	78
4.2	Interscale Correlations for Affective Quality Scales	81
4.3	Reliability: Intraclass Correlations and Kappa	82
4.4	Descriptive Statistics for Affective Quality Scales	83

4.5 Means on Teaching Tasks Ratings, by Subgroups Formed at Baseline 85

4.6 Means on Teaching Tasks Ratings, by 18-Month Follow-Up Characteristics 89

4.7 Impacts of New Chance on Teaching Tasks Ratings 95

4.8 Impacts of New Chance on Mother's Supportive Presence, by Subgroups Formed at Baseline 96

4.9 Impacts of New Chance on Mother's Harsh Treatment, by Subgroups Formed at Baseline 99

4.10 Impacts of New Chance on Child's Compliance, by Subgroups Formed at Baseline 102

4.11 Impacts of New Chance on Child's Affection to Mother, by Subgroups Formed at Baseline 105

4.12 Impacts of New Chance on Quality of Relationship, by Subgroups Formed at Baseline 108

5.1 Means on Maternal Book Reading Task Variables, by Subgroups Formed at Baseline 128

5.2 Global Rating Frequencies for Book Reading Quality 133

5.3 Means on Maternal Book Reading Task Variables, by 18-Month Follow-Up Characteristics 135

5.4 Impacts of New Chance on Maternal Book Reading Task Variables 141

5.5 Impacts of New Chance on Mother's Total Number of Utterances During Book Reading, by Subgroups Formed at Baseline 142

5.6 Impacts of New Chance on Mother's Number of Nonimmediate Utterances During Book Reading, by Subgroups Formed at Baseline 145

5.7 Impacts of New Chance on Mother's Percentage of Immediate Utterances During Book Reading, by Subgroups Formed at Baseline 147

5.8 Impacts of New Chance on Number of Discussion Topics During Book Reading, by Subgroups Formed at Baseline 150

5.9 Impacts of New Chance on Book Reading Quality, by Subgroups Formed at Baseline 152

5.10 Means on Wheels Task Variables, by Subgroups Formed at Baseline 157

5.11 Means on Wheels Task, by 18-Month Follow-Up Characteristics 160

5.12 Impacts of New Chance on Wheels Task Variables 164

5.13 Impacts of New Chance on Ease of Ideas (Wheels Task), by Subgroup at Baseline 166

6.1 Means on HOME-SF Scales, by Subgroups Formed at Baseline 178

6.2 Means on Maternal Report Parenting Scales, by Subgroups Formed at Baseline 181

6.3 Means on Parenting Time Use Measures, by Subgroups Formed at Baseline 184

6.4 Means on HOME-SF Scales, by 18-Month Follow-Up Characteristics 189

6.5 Means on Maternal Report Parenting Scales, by 18-Month Follow-Up Characteristics 193

6.6 Means on Parenting Time Use Measures, by 18-Month Follow-Up Characteristics 196

6.7 Impacts of New Chance on Survey Measures of Parenting at 18-Month Follow-Up 201

6.8 Impacts of New Chance on the HOME-SF Emotional Support Subscale, by Subgroups Formed at Baseline 202

6.9 Impacts of New Chance on the Maternal Warmth Scale, by Subgroups Formed at Baseline205 205

6.10 Impacts of New Chance on the Parenting Chore Time Measure, by Subgroups Formed at Baseline 208

7.1 Within-Group Analyses Means on Parenting Outcomes, by Program Participation 220

7.2 Comparing the Five Types of Parenting Measures Included in the New Chance Observational Study 225

7.3 Correlations Among Various Parenting Measures 233

7.4 Summary Information: Correlations of Observational and Interview-Based Measures of Parenting 235

7.5 Summary Statistics for Analyses Examining Sets of Parenting Variables as Predictors of Mother-Reported Behavior Problems Index 239

7.6 Summary Statistics for Analyses Examining Sets of Parenting Variables as Predictors of Mother-Reported Positive Behavior Index 240

7.7 Summary Statistics for Analyses Examining Sets of Parenting Variables as Predictors of Teacher-Reported Behavior Problems Index 241

7.8 Summary Statistics for Analyses Examining Sets of Parenting Variables as Predictors of Teacher-Reported Positive Behavior Index 242

7.9 Summary Statistics for Analyses Examining Sets of Parenting Variables as Predictors of Scores on Bracken Basic Concept Scale 243

8.1 Summary Statistics for Analyses Examining Sets of Variables as Predictors of Mother-Reported Behavior Problems Index 256

8.2 Summary Statistics for Analyses Examining Sets of Variables as Predictors of Mother-Reported Positive Behavior Index 257

8.3 Summary Statistics for Analyses Examining Sets of Variables as Predictors of Teacher-Reported Behavior Problems Index 258

8.4 Summary Statistics for Analyses Examining Sets of Variables as Predictors of Teacher-Reported Positive Behavior Index 259

8.5 Summary Statistics for Analyses Examining Sets of Variables as Predictors of Scores on Bracken Basic Concept Scale 260

10.1 Structural Typology of Measures in the New Chance Observational Study 294

12.1 Combinations of Person(s) Other Than Respondent and Focal Child Present During the Observational Session 312

12.2 Person(s) Other Than Respondent and Focal Child Present During the Observational Session 313

12.3 Interruptions and Distractions During the Observational Session 314

12.4 Seating Arrangements of Mother and Child During Observational Session 315

12.5 Types of Deviations in Session Administration Noted in Coders' Comments, for the Book Reading and Wheels Tasks 319

12.6 Severity of Deviations in Session Administration Noted in Coders' Comments, for the Book Reading and Wheels Tasks 321

12.7 Distribution of Completed Observational Sessions, by Summary Severity of Deviations Scores for the Book Reading and Wheels Tasks 323

12.8 Distribution of Interviewers and Observational Sessions, by Interviewer's Average Summary Severity of Deviations Score 324

12.9 Average Summary Severity of Deviations Scores of Interviewers, by Number of Sessions Completed 325

12.10 Maternal Self-Reports of the Subjective Experience of the Observational Session 328

12.11 Indicators of Mothers' and Children's Subjective Experience of the Observational Session 331

12.12 Correlations Among Subjective Experience Measures 334

12.13 Correlations of Child's Age with Subjective Experience Measures 337

12.14 Predictors of Mothers' and Children's Subjective Well-Being During Observational Session 339

A.1 Regression Coefficients from a Regression of the Experimental Dummy on Selected Baseline Characteristics 354

B.1 Observational Study Respondents and Initial Pool of Respondents in the Seven Study Sites 355

C.1 Regression Coefficients for Analyses Examining Multiple Influences on Mother-Reported Behavior Problems Index 356

C.2 Regression Coefficients for Analyses Examining Multiple Influences on Mother-Reported Positive Behavior Index 358

C.3 Regression Coefficients for Analyses Examining Multiple Influences on Teacher-Reported Behavior Problems Index 360

C.4 Regression Coefficients for Analyses Examining Multiple Influences on Teacher-Reported Positive Behavior Index 362

C.5 Regression Coefficients for Analyses Examining Multiple Influences on the Bracken Basic Concept Scale 364

Foreword

This is a report on the New Chance Observational Study – an in-depth examination of parenting behavior and its relationship to children's development in a subset of the families participating in the New Chance Demonstration.

New Chance was a national research and demonstration project that provided comprehensive education, training, and other services intended to improve the prospects and well-being of low-income mothers and their children. The program's eligibility criteria were designed to assure that the research sample represented families central to the welfare reform debates of the past 10 years: families headed by young mothers who had their first child as teenagers, were high school dropouts, and were receiving Aid to Families with Dependent Children. (AFDC was the main cash welfare program until the 1996 federal welfare legislation replaced it with Temporary Assistance for Needy Families, or TANF.)

One of New Chance's distinguishing features was its explicit two-generational focus on both mothers and children. Many of its services were meant to help the mothers prepare for, get, and hold onto jobs so that they could become economically self-sufficient and leave welfare. In designing the demonstration, MDRC believed that such changes, if they occurred, would potentially improve developmental outcomes for children. But the aim was to shape the children's development more directly. Therefore, the demonstration included parenting education, access to pediatric health services, and an attempt to develop and encourage the use of good-quality child care.

The original research plan for evaluating the New Chance Program recognized the importance of assessing its two-generational nature. Thus, it included several measures of parenting behavior and participation by the mother in the child-related components of New Chance (e.g., parenting education), along with various measures of the children's development. Sources for the latter included surveys of parents, reports from teachers (for children in the research sample who were in formal child care or school settings), and a general measure of school readiness.

While the initial design of the New Chance Program and its research plan went beyond the practices typical of large-scale field studies of this type, they created the opportunity to push even further. Thus, MDRC formed an interdisciplinary "observational studies team" that designed and implemented the New Chance Observational Study and a companion study embedded within the evaluation of the Job Opportunities and Basic Skills Training (JOBS) Program (the welfare program authorized under the legislation that preceded the legislation establishing TANF).

Each of the observational studies makes a new and unique contribution regarding how best to measure parenting and child outcomes in a survey context. This is a question important to anyone trying to understand the effects on children of welfare reforms or similar initiatives. Much of the existing information about how children develop comes from the university-based laboratories of developmental psychologists, yet social policies and programs play out in the everyday lives of parents and children. The observational studies transported university-based techniques into the field, gathering data via videotape and audiotape in the participants' homes. In large part, this translation of methods and materials succeeded, and it yielded rich information about the strengths and limits of using regular survey interviewers to go well beyond their typical interviewing tasks. The information on all this methodological work is found in the second part of this monograph.

The incorporation of these new techniques and measures into the New Chance Demonstration also allowed us to deepen our answers to certain questions. For example, using the originally planned survey measures, MDRC staff have examined questions about the effects of New Chance on parenting, and the relationships between those effects and child outcomes. But these analyses acknowledged the limits of data on parenting or child development that come from parental self-report and observations by survey interviewers. By including the measures derived from videotapes and audiotapes of the observational study sessions, the observational studies team was able to more fully explore the effects of New Chance on parenting, the relationships between program participation and these effects, and the role of the parenting effects on child outcomes.

Interestingly, the New Chance program *did* have positive effects on parenting. These appeared on both survey interviews and observational measures. However, other influences such as maternal psychological well-being, including stress and depression, combined with the effects of parenting behavior on children. Thus, the positive parenting differences were not sufficient to bring about effects on child outcomes, no matter how the parenting and child outcomes were measured.

While the observational study data were consistent with the survey data and thus did not change our judgments about the effects of the New Chance Program, they were promising in other ways. The observational data appear to have tapped a wider range of parent and child behaviors than is possible in the survey context. For example, the survey measures did not detect program impacts in the area of cognitive stimulation, while the more sensitive and qualitative observational measures did. The observational measures also added to our ability to understand why some children are doing better than others. Such questions and analyses are the subject of the first part of the monograph.

Interdisciplinary and pathbreaking, the observational study represented a collaboration among a diverse team. The editors of the monograph, Martha Zaslow and Carolyn Eldred, have generously described the roles of all team members except themselves in the Acknowledgments. Each of the editors has given an extraordinary amount of time to this effort over the last six years, and their contributions cannot go unnoted.

Martha Zaslow has been involved in all aspects of this work. Together with colleagues at Child Trends, she analyzed the interview-based measures of parenting, examined the relationships across different parenting measures, and conducted analyses regarding the role of parenting and other factors in predicting child outcomes. She also wrote or contributed to several chapters in the first part of the monograph and reviewed all of the manuscript several times.

Carolyn Eldred began the effort as the project director for the New Chance surveys at the Institute for Survey Research at Temple University. In that role, she worked with MDRC to hone the overall study design, thought through the myriad issues involved in conducting and recording the observational sessions, adapted the laboratory protocols for survey administration, trained the field staff, and directed the data collection effort. Subsequently, as a consultant to MDRC, she undertook the methodological reflection and analysis in the second part of the report, wrote the chapters on methodological issues, coauthored the chapter on study design in the first part of the report, and reviewed the entire monograph several times. Both editors have been "first among equals" on this team.

The study owes a great debt to the energy and good efforts of the mothers who invited us into their homes, diligently worked through the tasks with their children, and shared their thoughts in the interviews. Above all else, the research team hopes that this work is up to the trust shown by these families. They have been open and generous; this monograph is dedicated to improving their lives and the lives of others in similar circumstances.

Finally, we are grateful to the funders of the observational study: the Foundation for Child Development, the William T. Grant Foundation, the Smith Richardson Foundation, and the National Institute of Child Health and Human Development. Their commitment made the project possible.

Robert C. Granger
Project Director

Acknowledgments

This monograph reflects the contributions of many individuals and organizations. The editors gratefully acknowledge their work, insight, and encouragement.

The monograph results, first, from the willingness of the young women in the New Chance Observational Study sample to open their homes and lives for the purposes of this research. This study could not have taken place if they had not shared their experiences, thoughts, and time with us.

We are grateful to the project funders, who provided both the financial support that made the research possible and much thoughtful input throughout the course of the research. Sheila Smith of the Foundation for Child Development and Lonnie Sherrod of the William T. Grant Foundation helped launch the project and have followed it closely in every stage of its development, most recently serving as external reviewers of an earlier draft of this monograph. Grant 5R01 HD31056-03 from the National Institute of Child Health and Human Development provided the funds for the methodological analyses reported here, which look at the relationship between interview-based and observational measures of parenting, and their relative importance in predicting child outcomes. The project was also supported by the Smith Richardson Foundation.

The transformation of data and analyses into this final product owes much to the insightful comments of those who reviewed drafts of the monograph, including: Jeanne Brooks-Gunn of Teachers College at Columbia University, Deborah Coates of the City University of New York, E. Mavis Hetherington of the University of Virginia, Aletha C. Huston of the University of Texas, and Judith M. Tanur of the State University of New York at Stony Brook.

The research could not have been carried out without the willingness of the Institute for Survey Research (ISR) at Temple University and its staff to take on the challenge of adapting the laboratory protocols for field use and collecting the data. Ellin Spector provided ongoing input to the work, while Marie DiCamillo and, later, Sonia Lawson supervised the data collection, with the assistance of Linda DeWild. The late Marcia Murphy oversaw creation of the data file and provided valuable assistance with sample definition and selection. Peter Buffum assisted with overall project direction. Richard Luker, then of Temple University's Radio, Television, and Film Department, aided ISR's work by designing the approach for recording the sessions with the mothers and their children and dealing with myriad technical decisions. Douglas Keith and Ash Meer oversaw the videography. Especially critical to the success of this endeavor were the survey interviewers and the videographers, who stretched their skills in new ways to collect the study data.

The editors are extremely grateful for the process of mutual and interdisciplinary education that occurred over a period of years across the different groups that constituted the New Chance Observational Study research team. In particular, we thank:

Robert Granger at the Manpower Demonstration Research Corporation (MDRC). As project director of the New Chance Demonstration at MDRC, Robert Granger took on the leadership and coordination of the New Chance Observational Study as well. He assembled the consortium of researchers who contributed to the observational study, provided ongoing guidance regarding all aspects of the research, and helped guide this monograph to fruition.

Other colleagues and consultants at MDRC. Denise Polit contributed to the study design and monitored the fielding of the survey work for the project. Hans Bos created the final data file and documentation, contributed to virtually all data analysis decisions, oversaw the creation of the final data file, and was the senior reviewer for all aspects of the data analysis. Janet Quint provided invaluable advice about and understanding of the teen mothers' experiences, which aided in designing and carrying out the research. She also helped us conceptualize the structure of the monograph. Vicki Hong processed much of the data for questions about the members of the research sample, and performed analyses based upon more specialized requests from the Child Trends team and Carolyn Eldred. Christine Schwartz was responsible for additional data analysis. Rachel Cytron, assisted by Karen Trister, attended to the innumerable details of report and table production. She also coordinated the final production details among authors and editors. Copyediting was handled deftly by Sylvia Newman, and Judy Greissman edited the Executive Summary. Patt Pontevolpe and Stephanie Cowell were responsible for word processing.

Byron Egeland, Nancy Weinfield, and John Ogawa of the University of Minnesota. Byron Egeland provided the protocol and props for several of the mother-child interaction tasks, and helped revise the protocol for survey administration. The University of Minnesota coding team included: Byron Egeland, Nancy Weinfield, Marnie Hiester, Catherine Lawrence, Katherine Hennighausen, Kathy Johnston, Susan Pierce, and Jenifer Powell. This team developed and refined the rating system for affective quality of interaction, and then coded the project videotapes. Coding was coordinated by Marnie Hiester, Catherine Lawrence, and Nancy Weinfield, successively. John Ogawa advised the coding team on statistical and methodological issues, created the data file for the Minnesota data, and conducted all the analyses on the affective quality data. Nancy Weinfield, Byron Egeland, and John Ogawa wrote up the results concerning this aspect of mother-child interactions, collaborated on the joint team chapter focusing on the prediction of child outcomes, and reviewed all other chapters. This study would not have been possible without the assistance of Mervyn Bergman who created many of the props (e.g., the blocks and the sorting boards).

Catherine Snow, Jeanne De Temple, Patton Tabors, and Brenda Kurland of Harvard University also collaborated on shaping the protocol for the mother-child interaction tasks. Jeanne DeTemple, Eliza Whitbeck, Claudia Cooper, and Jane Herman transcribed the mother-child interactions from two of the tasks, coded the literacy-related mother-child interaction from the transcripts and videotapes, and provided the commentary on the sessions that underlies the methodological analyses in the second part of this report. They analyzed and wrote up the results on mother-child interaction related to the emergence of literacy, collaborated on the chapter that looks at the predictors of child outcomes, and provided feedback on all the chapters.

M. Robin Dion of Child Trends not only managed Child Trends' work on data analysis, table creation, and many writing tasks for the monograph, but also served as coordinator for several tasks conducted in parallel across all the research groups, including subgroup definition and programming for descriptive and impact analyses. She also carried out the interviews with the field staff that helped us all get a deeper sense of the parenting education component of the New Chance Program, carried out a number of the analyses reported on here, served as lead author on one of the chapters, and contributed to the writing and revising of further chapters.

Kristin Moore, Carol Emig, Christopher Botsko, Carrie Mariner, and Tawanda Greer, also of Child Trends, reviewed drafts of the monograph. Carol Emig is leading the team in communicating the findings of this monograph beyond the research community. Tawanda Greer assisted in data analysis and table creation, as did Fanette Jones on table creation.

Donna Ruane Morrison of Georgetown University advised on many data and analysis issues, leading the analysis and writing up of findings on the interview-based measures of parenting, and reviewing and helping in the revision of the monograph as a whole.

Deborah L. Coates of the City University of New York, Vonnie C. McLoyd of Duke University, and Craig and Sharon Ramey of the Civitan International Research Center of the University of Alabama at Birmingham consulted throughout the course of the project. We are grateful for the advisory role they played in this research.

The Editors

Executive Summary

Martha J. Zaslow and Carolyn A. Eldred

Part I: The New Chance Observational Study

I. The New Chance Demonstration

The New Chance Observational Study — the subject of this monograph — is an in-depth examination of parenting behavior in 290 of the 2,322 families studied in the New Chance Demonstration, a national research and demonstration program operated between 1989 and 1992 at 16 locations in 10 states. The demonstration tested a program model intended to improve the economic prospects and overall well-being of low-income young mothers (aged 16 to 22) and their children through a comprehensive and intensive set of services. It was developed by the Manpower Demonstration Research Corporation (MDRC) and supported by a broad consortium of public and private funders.

New Chance was directed at families central to the welfare reform debates that culminated in the Personal Responsibility and Work Opportunity Reconciliation Act of 1996 — families headed by young mothers who gave birth during their teenage years and were receiving Aid to Families with Dependent Children (AFDC, the main cash welfare program).[1] More specifically, New Chance focused on those who were especially disadvantaged because they were high school dropouts; as a group, they and their children are at high risk of long-term welfare receipt and economic hardship.

The New Chance Program sought to help the young mothers (who, for the most part, volunteered for the program) to acquire educational and vocational credentials and skills so that they could find and keep jobs offering opportunities for advancement and reduce, and eventually eliminate, their use of welfare. It also sought to motivate and assist participants to postpone additional childbearing and to become better parents. Because New Chance focused on young children as well as their mothers, it sought to further the cognitive, social, and emotional development as well as the health of participants' children. Child care was provided at no cost to the parents, on site in most places, and the program facilitated access to health services for both mothers and children. The program was intended to be intensive (four to five days a week for up to 18 months), though in practice attendance was of much shorter term and often irregular.[2]

---

[1] The law, which replaced the entitlement to AFDC with a block grant to states, requires unmarried minor parents caring for a child to live with an adult, or in a setting supervised by an adult, and to participate in education or training activities in order to receive welfare assistance. The law also calls for the U.S. Department of Health and Human Services to develop and implement a strategy to prevent nonmarital births to teens and to assure that at least 25 percent of communities have teen pregnancy prevention programs.

[2] The New Chance Program and its implementation are described in detail in Janet C. Quint, Barbara Fink, and Sharon Rowser, *New Chance: Implementing a Comprehensive Program for Disadvantaged Young Mothers and*

(continued)

II. Parenting Behavior and the Two-Generational Character of New Chance

New Chance was one of only a few interventions for families in poverty that took a two-generational approach, seeking to improve the outcomes for both the young mothers, who faced multiple difficult life circumstances, and their children. For this reason, parenting behavior took on particular importance. If the New Chance Program improved the quality of the mother-child relationship, it would signify improved functioning in an important area of the mothers' lives. It would also have the potential of diminishing risk and improving developmental outcomes for the children. Research shows that the children of young single mothers in poverty begin to show higher levels of behavior problems as early as the preschool years. Later in development, they show difficulties in school progress and achievement.

New Chance sought to enhance parenting behavior most directly through its parenting classes, which provided information on children's developmental stages, activities and materials to enhance children's cognitive development, and developmentally appropriate strategies for shaping child behavior. Parenting classes, which were scheduled for about two hours a week, balanced open discussion of issues of concern to the mothers and more formal presentation of specific information.

Other aspects of New Chance also held the potential for enhancing parenting behavior. Life skills training (like parenting classes, scheduled for about two hours a week) focused on improving the mothers' skills in communicating with significant people in their lives, including children. Adult basic education (including preparation for the General Educational Development, or GED, test) and job skills training classes provided stimulation to the mothers, which in turn could be reflected in how they talked with, read to, and played with their children. Group and individual counseling addressed problems emerging in the mothers' lives, including problems with children. Family planning classes stressed the importance of providing enough time and attention to each child in the family. The child care that the children participated in provided a context for the young mothers to observe care providers engaging in stimulating and supportive interactions with children — behavior they might then imitate. Participation in the program as a whole could increase the mothers' sense of social support, which in turn could enhance parenting behavior.

III. The New Chance Observational Study: A Study Embedded in the New Chance Evaluation

A. Evaluating the New Chance Program

Central to the New Chance Demonstration was a rigorous evaluation of the program's effectiveness. For this purpose, 2,322 young women who applied to New Chance were randomly assigned to either an *experimental* group (who were allowed to enroll in New Chance) or a *control* group (who did not have access to services provided through New Chance, but many of whom found some alternative services in their communities). To determine the effectiveness of

---

Their Children (New York: MDRC, 1991).

the New Chance "package" of services, compared with these alternative services, *differences* in outcomes between the two groups of women and their children (often referred to in this kind of research as the *impacts* of the program) were examined through structured survey interviews administered approximately 18 and 42 months after each young woman entered the research sample. The findings from the demonstration have been presented in a series of MDRC reports concluding with the final report, released in 1997.[3]

B. The Purposes of the Observational Study

Because of the importance of parenting behavior in the context of a program focusing on outcomes for two generations, a special study was undertaken, using videotaped observations of mother-child interaction among a subset of 290 families in the New Chance Demonstration. The work of a collaborative and interdisciplinary research team (described in the Foreword to this monograph), the observational study was funded by grants from the Foundation for Child Development, the William T. Grant Foundation, and the Smith Richardson Foundation, with specific analyses focusing on methodological issues funded by the National Institute of Child Health and Human Development.

The observational data were collected during a special session following the 18-month follow-up survey for the full evaluation, on average 21 months after each member of the observational study sample had enrolled in the demonstration. The videotapes of mother-child interaction were then coded under rigorous conditions in university laboratories. The purposes of the study were to (1) describe parenting behavior in this sample, especially the affective quality of mother-child interaction and the aspects of interaction that are related to the emergence of literacy in children; (2) examine in greater depth the program's impacts on parenting behavior; (3) explore the role of parenting behavior in shaping the outcomes for children; and (4) assess the added value of using measures of parenting based on direct observation in addition to the evaluation's survey interviews.

More specifically the study asked:

- What background characteristics of the families are most closely related to parenting behavior for the families in the sample?
- Based on sensitive and detailed measures of parenting used in the observational study, did the New Chance Program have positive impacts on parenting behavior and, if so, which program components contributed?
- What role did parenting behavior play in shaping the development of the children in the sample? What was the role of such other important influences as the mothers' psychological well-being and the family's economic resources?

The study also asked how best to measure parenting behavior in a program evaluation that focuses on two generations. More precisely:

---

[3] Janet C. Quint, Johannes M. Bos, and Denise F. Polit, *New Chance: Final Report on a Comprehensive Program for Young Mothers in Poverty and Their Children* (New York: MDRC, 1997).

- Do measures of mother-child interaction based on direct observation substantially improve the quality of information about parenting beyond what is available through measures collected in the context of survey interviews? Part II of this Executive Summary (and of the monograph) discusses the methodological issues in more detail.

C. The Sample

The sample for the New Chance Observational Study was chosen from the seven New Chance sites that had the largest number of families who met the study's criteria and could potentially participate. Specifically, the study chose families with a "focal child" between 30 and 60 months old, an age range considered appropriate for the study's procedures. (The "focal child" was the child in each family who was the focus of the interviews and assessments in the full evaluation's impact research.) In addition, the observational study was limited to African American and white families, rather than also including Hispanic families, because the researchers did not consider it possible to give full consideration to variation in parenting behavior that might be associated with cultural background. To study how parenting behavior related to other important aspects of the families' lives, the sample was also limited to families who had completed the evaluation's 18-month follow-up interview. Finally, the study was restricted to families for whom the observational study session could be scheduled within a similar time frame relative to random assignment — no more than four months after each mother's 18-month follow-up interview. The eligibility criteria for this study were more restrictive than the criteria for the overall New Chance Evaluation, and the sites used in the observational study were not selected randomly from all the New Chance sites. Therefore, the findings from the observation study should not be seen as generalizing to the full New Chance sample. The larger evaluation sample, for example, included families with children in a broader age range, families from additional sites, and families of Hispanic origin.

The sites from which the families were chosen were the Bronx, Detroit, Harlem, Lexington (Kentucky), Philadelphia, Pittsburgh, and Portland (Oregon). Within each of these sites, families who met the study's eligibility criteria were contacted in the order in which they had enrolled in the demonstration.

The goal was to conduct observational sessions with about 300 families — a sample size large enough to detect program impacts on the observational measures and to permit an examination of how parenting behavior was related to the mothers' education, age at first birth, years receiving public assistance, and other important characteristics, as well as child outcomes. Of the New Chance sample members who met the criteria for inclusion in the observational study, 79 percent participated, yielding usable videotapes from 290 families. Approximately 84 percent were African American and 16 percent were white. Among the focal children, 148 were boys and 142 were girls. One hundred and eighty-four of the mothers were members of the demonstration's experimental group and 106 were from the control group. The families in the two groups did not differ significantly in their baseline characteristics, so group differences found at the time of the observational study can be attributed to the experiences of the families subsequent to random assignment. Nor was there evidence of systematic differences between the families identified as eligible to participate in the observational study who did and did not participate.

D. An Overview of the Study's Procedures

The visit to each family participating in the observational study (lasting, on average, about an hour) was conducted by a two-member team — a survey interviewer, who briefly interviewed the mother and guided her through a series of interactive tasks with her child, and a videographer, who taped the mother and focal child as they carried out the interactive tasks. The interviewers, who were accustomed to conducting traditional survey interviews and were already working on the New Chance 18-month survey, received specialized training to follow a structured script and to explain and administer the tasks to the mothers with minimal interference in the mother-child interaction.

The observational session, or extra visit carried out for the New Chance Observational Study, started with the interviewer explaining the procedures to the mother and obtaining her informed consent. The interviewer then administered the first half of a brief interview with the mother, in which she was asked to describe, in half-hour intervals, the activities that she and her child had engaged in during the previous weekday. Beginning the session in this way helped the interviewer to establish rapport with the mother, gave the videographer time to set up the equipment, and provided further information about the mother-child relationship.

Following a script, the interviewer then described each of the interactive tasks to the mother and confirmed that she understood how to carry out each task. After these initial instructions, the child was invited to join the mother and interviewer, and the mother was asked to guide her child through each of the interactive tasks. The script called for the interviewer to remind the mother briefly about each task as she presented the props for that task. Interviewers were instructed to then let the mother interact with her child as she chose, without interference.

The tasks, devised by university-based researchers of children's development, had been used in previous studies of mother-child interaction, including studies of low-income families. The six tasks were:

- **Book reading.** The mother was asked to read and discuss a children's book, *The Very Hungry Caterpillar* (by Eric Carle), with her child the way she would usually do so;
- **Blocks.** The mother was asked to try to get her child to match the shape of a larger block by using combinations of smaller blocks;
- **Wheels.** The mother tried to get her child to name as many objects with wheels as he or she could within the time allotted;
- **Sorting.** The mother asked her child to place plastic chips of different shapes and colors in the empty squares on a board according to the shapes of chips glued onto the board in a row at the top, and the colors of chips glued on in a column at the left;
- **Etch-a-Sketch.** The mother tried to get her child to use the knobs on an Etch-a-Sketch board to draw a line tracing a maze that had been drawn on the screen;

- **Gift.** The mother was presented with a wrapped gift — a kaleidoscope — to give to the child, and the mother and child then spent a few minutes opening and playing with the gift.

Each of these tasks had either been used by members of the research team in previous studies or was a modification of a task used before. The tasks were chosen because they yield forms of mother-child interaction that predict social behavior and academic achievement in school.

The observational session concluded with the interviewer administering the final portion of the brief interview to the mother, which included questions about use of child care for the focal child; the mother's participation in educational, training, and employment activities; and the family's residential situation. The interviewer also completed ratings about the home environment and about the observational session (for example, whether there had been others in the home during the session and whether the session had been interrupted), and the mother completed a brief "self-administered questionnaire" with items concerning her subjective sense of well-being, perceptions of the focal child, and reactions to the mother-child interaction tasks.

The initial instructions to the mother as well as the series of mother-child interaction tasks (but not the interview segments of the observational session) were videotaped. The videotapes of mother-child interaction were then coded in two independent research laboratories. The affective quality of mother-child interaction was rated by a team of researchers at the University of Minnesota. Ratings of mother-child interactions related to the emergence of literacy were made by a team of researchers at Harvard University, based on transcriptions of verbal interactions during the book reading and wheels tasks and on the interactive behavior from the videotapes for these tasks.

E. Parent-Child Data from the Full New Chance Evaluation

The families in the New Chance Observational Study also participated in the data collection that was part of the full New Chance Demonstration. Thus, there is information on the families prior to their being randomly assigned to the evaluation's experimental or control groups ("baseline data"), before any program effects could have occurred. There are also data from the 18- and 42-month follow-up surveys, both of which included measures of the mothers' psychological well-being; educational attainment, employment, earnings, and welfare receipt; residential situation; use of child care for the focal child; and fertility.

The surveys from the full evaluation also included interview-based measures of parenting. The analyses reported on in this monograph focus on parenting measures from the 18-month follow-up survey, because these were collected close in time to the observational measures. The 18-month follow-up included three parenting scales based on maternal report: Warmth, Control, and Parenting Stress. The 18-month follow-up also included a measure of the emotional support and cognitive stimulation available to the child in the home environment, based on a combination of questions asked of the mother and ratings completed by the interviewer. This measure, the Home Observation for Measurement of the Environment-Short Form (or HOME-SF), provided a total score as well as Emotional Support, Cognitive Stimulation, Harsh Discipline, and Physical Environment subscale scores.

Table 1 lists the parenting measures available for the sample of the New Chance Observational Study and the source of each measure. It is important to note that the parenting measures included in the New Chance Observational Study rely on several different informants (mothers, interviewers, and coders of the observational session videotapes). Thus, we are not confined to one data source in examining the New Chance Program's possible impacts on parenting behavior. Each of the measures (along with information about its previous use and psychometric properties) is described in detail in the monograph.

The focal children's development was assessed as part of the 42-month follow-up. Direct assessments of their cognitive development were carried out using the School Readiness Component of the Bracken Basic Concept Scale, which assesses children's knowledge of such concepts as colors, letters, numbers, shapes, counting, and making comparisons. In addition, mothers rated their children's behavior problems (using the Behavior Problems Index) and positive social behaviors (using the Positive Behavior Index) and responded to questions concerning their children's health. The Behavior Problems Index provides a total score and subscale scores for behaviors that reflect antisocial, anxious/depressed, headstrong, hyperactive, dependent, and peer conflict/withdrawal behaviors. The Positive Behavior Index provides a total score and subscale scores for compliance, social competence, and autonomy.

For those children already in a formal child care situation or school setting, teachers were asked to complete a survey, in which they rated children's academic and behavioral adjustment to school. They also completed the Behavior Problems Index and the Positive Behavior Index.

IV. The Findings in Brief

- **Mothers in the New Chance Observational Study sample were at particularly high risk in terms of parenting behavior.**

Some of the parenting measures used in the New Chance Observational Study had been used in other studies of high-risk families. On these measures, New Chance mothers had less positive scores than mothers in the other high-risk samples. When reading to their children, New Chance mothers showed lower frequencies of "Nonimmediate Utterances," that is, speech that extends beyond the information in the book (the specific words in the text and the pictures shown) to make connections with other experiences and information. Previous research shows Nonimmediate Utterances to be important in laying the groundwork for later literacy in children. Mothers in the New Chance Observational Study also expressed hostility to their children more often than the mothers in another high-risk sample. Harsh mother-child interaction is important to development, predicting less positive adjustment in children.

- **Despite their similar economic circumstances and backgrounds, the mothers in the New Chance Observational Study showed variation in their parenting behaviors, and parenting behaviors were meaningfully related to the mothers' background characteristics.**

Among mothers in the New Chance Observational Study sample, more emotionally supportive and cognitively stimulating parenting behaviors were associated with higher maternal

Table 1

Parenting Measures
Included in the New Chance Observational Study

Type of Parenting Measure	Variables	Time of Collection in New Chance Study: Observational Session	Time of Collection in New Chance Study: 18-Month Interview	Time of Collection in New Chance Study: 21-Month Interview	Chapter
Observational Measures of Mother-Child Interaction Related to Literacy	*Book Reading Task* Total Number of Utterances Number of Nonimmediate Utterances Percentage of Immediate Utterances Number of Discussion Topics Book Reading Quality *Wheels Task* Objects Named Objects/Elicitations Mother's Ease of Ideas	Derived from two tasks only: book reading task and wheels task			5
Observational Measures of Affective Quality of Mother-Child Interaction	*Ratings of Mother* Supportive Presence Intrusiveness Hostility Quality of Instruction Confidence Harsh Treatment *Ratings of Child* Persistence Enthusiasm Negativity Compliance Experience of Session Affection to Mother Avoidance of Mother *Ratings of Dyadic Behavior* Quality of Relationship Boundary Dissolution	Derived from full videotape			4
Combination of Interviewer Ratings and Maternal Report Home Observation for Measurement of the Environment Short Form (HOME-SF)	Emotional Support Cognitive Stimulation Physical Environment Harsh Discipline HOME Total		X		6
Maternal Report Scales	Warmth		X		6
	Control		X		
	Stress		X		
	Overall Parenting Time			X	
	Parenting Chore Time			X	

literacy, more educational attainment, better maternal psychological well-being, greater social support, and the child's having participated in child care.

- **The New Chance Program was able to bring about positive changes in parenting behavior, even in a population burdened by economic stress and other serious difficulties.**

Positive program impacts, although modest in magnitude, were seen in both the affective quality of mother-child interaction and the cognitive stimulation that the mothers provided. In terms of the *affective quality of interaction*, mothers in the experimental group had significantly lower scores on the Harsh Treatment measure, higher scores on the HOME-SF Emotional Support subscale, and higher scores on the Maternal Warmth scale. In terms of *cognitive stimulation*, mothers in the experimental group received higher scores on the Book Reading Quality measure. Positive program impacts were also found on the HOME-SF total score and on time use measures devoted to parenting (both Overall Parenting Time and Parenting Chore Time — time spent engaged in such chores as feeding and bathing children).

- **Positive program impacts were found across parenting measures obtained in several different ways and from different informants.**

Differences were found for parenting measures based on direct observation of mother-child interaction (Harsh Treatment and Book Reading Quality), measures based entirely on maternal report (Warmth, Overall Parenting Time, and Parenting Chore Time), and measures that rely on a combination of maternal report and interviewer ratings (the HOME-SF total score and Emotional Support subscale).

- **Parenting behavior was an important predictor of specific child outcomes in this sample, as were variables reflecting maternal psychological well-being and the families' larger social context.**

Positive implications of supportive and stimulating parenting behavior combined with the negative implications of maternal psychological distress and stress in the larger social context to shape the developmental outcomes of children in this sample.

- **Modest improvements in parenting behavior, in this context, did not suffice to bring about positive program impacts on child outcomes.**

Given the high levels of maternal psychological distress and stress in this sample, bringing about positive impacts on child outcomes might have required (1) a more intensive "dosage" of parenting education classes or other program components that enhance parenting behavior, (2) sufficiently intense program components directly focusing on the mothers' psychological well-being and living situations, and/or (3) high-quality child care, with children participating over a sustained period of time.

- **Observational measures add to the understanding of parenting behavior within this sample in multiple ways.**

Observational measures increase one's confidence in findings on program impacts because the coding is carried out in an extremely rigorous way. They also tap into certain behaviors that are important to children's development, but that mothers may not be able to report on (such as their use of Nonimmediate Utterances during book reading). Also, the findings indicate that observational measures have advantages over interview-based measures for predicting variation in children's scores on specific child outcomes. For example, consideration of the observational measures of mother-child interaction significantly improved prediction of the Behavior Problems Index (as reported on by the mother) even when parenting measures based on interviews had already been taken into account. In short, observational measures provide valuable information that can enrich the evaluation of programs such as New Chance.

V. Parenting Behavior in This Sample Relative to Other Samples

The observational study interaction tasks, and the coding of behavior from those tasks, grew out of previous research in the laboratories of Catherine Snow, Jeanne De Temple, and their colleagues at Harvard University, and Byron Egeland, Nancy Weinfield, John Ogawa, and their colleagues at the University of Minnesota. Since the measures have been refined and improved over time, it is generally not possible to compare directly the findings from the New Chance Observational Study with those from previous studies. For a few specific measures, however, modifications in rating scales have not occurred. When we look at findings for those measures that can be compared across studies, we see that New Chance Observational Study families are at particularly high risk in terms of parenting behavior.

One key measure of mother-child interaction related to the development of children's literacy is the proportion of maternal talk during the book reading task that involves Nonimmediate Utterances, that is, connecting the story and pictures to other events, people, and objects. Such talk also involves making predictions, asking for inferences, and providing explanations. Previous research using this measure in the Home-School Study of Language and Literacy Development, a longitudinal study of low-income mothers and their children, found maternal use of Nonimmediate Utterances to be related to important child outcomes in the early years of elementary school, especially the children's use of language and their literacy skills.

Nonimmediate Utterances in the context of a book reading task constituted 10.5 percent of mothers' utterances in the Home-School Study of Language and Literacy Development, but only 3.5 percent of talk among New Chance Observational Study mothers. The researchers note that "Since we have found that maternal use of Nonimmediate Talk relates to later child outcomes . . . the very low proportion of Nonimmediate Talk produced by the New Chance mothers is troubling."

The ratings scales of the affective quality of mother-child interaction used in the New Chance Observational Study are adaptations of scales developed in the Minnesota Mother-Child Project, a longitudinal study of high-risk mothers and children. The Hostility rating scale has not been substantially modified for the present study, and results can be compared across studies. Ratings of 5 or above indicate that a mother is more hostile than not in interacting with her child during the course of the mother-child tasks. Seven percent of the New Chance Observational

Study sample scored 5 or above compared with 2.7 percent on the Minnesota Mother-Child Project sample.

These findings suggest that the New Chance Observational Study sample is at greater risk in terms of parenting behavior than the previously studied high-risk samples, underscoring the importance of attempting to enhance parenting behavior in the population of young single mothers in poverty.

VI. Variation in Parenting Behavior in Light of Maternal and Family Characteristics

Family characteristics were significantly linked with parenting behaviors, and associations were more consistent with 18-month than baseline variables, suggesting that the more current family context is of greater importance to parenting behaviors. These characteristics included social support; measures of the mother's psychological well-being, residence pattern, maternal education, and literacy; and the child's participation in child care during the initial 18-month follow-up period.

For example, mothers who reported more sources of social support at the time of the 18-month follow-up interview had more positive relations with their children in terms of the observed affective quality of mother-child interaction, the observed literacy-related aspects of interaction, and the harshness of discipline as measured by the HOME-SF. Mothers at high risk for depression at the 18-month follow-up had lower scores on all but one of the HOME-SF subscales, had lower observed Book Reading Quality scores, reported more parenting stress, and described themselves as using more controlling disciplinary practices.

VII. Program Impacts on Parenting Behavior

Significant positive program impacts on parenting behavior were found on a range of parenting measures:

- Measures based on direct observation of mother-child interaction:

    Harsh Treatment

    Book Reading Quality

- Measures based on maternal report alone:

    Maternal Warmth

    Overall Parenting Time

    Parenting Chore Time

- Measures based on a combination of maternal report and interviewer ratings:

    HOME-SF total score

HOME-SF Emotional Support subscale

Positive program impacts occur in aspects of parenting that have previously been identified as particularly important to the development of children in poverty. For example, harsh parent-child interaction has been found to occur with greater frequency among families experiencing economic hardship. Harsh parent-child interaction, in turn, is a key contributor to the less favorable adjustment of children in poverty.

In general, the findings indicate that the New Chance Program had positive impacts on parenting behaviors important to development. There are three caveats, however. First, there were differences on only a minority of the parenting measures examined. Second, all but one of the program impacts, that on the HOME-SF Emotional Support subscale, were small in magnitude ("effect size"). Finally, a single significant finding ran counter to this pattern of positive impacts: Mothers in the experimental group received lower Ease of Ideas scores, observed during the mother-child task that called for eliciting from the child the names of objects that have wheels. That is, mothers in the experimental group were observed to be less facile in coming up with hints and clues for the child. This difference, however, was found to be attributable entirely to a larger number of mothers within the experimental group who did not grasp the goal of this task at all. Thus, this group difference might just as easily be interpreted as an indication of task or test anxiety, or of problems with the interviewers' explanations of this task, than as a reflection on the quality of the mothers' parenting behavior.

The research team concluded that the New Chance Program had positive, albeit modest, impacts on parenting behavior.

VIII. Program Components That Contributed to Positive Program Impacts on Parenting Behavior

Among mothers in the experimental group within the New Chance Observational Study sample, greater participation in parenting education classes was related to more positive parenting behavior. Interestingly, however, this pattern was also found for participation in *other* New Chance Program components.

We looked at four aspects of experimental group mothers' participation in New Chance: (1) participation in parenting education classes, (2) participation in a broader set of program activities that addressed parenting behavior in some way, (3) participation in human capital development components of the program (that is, components intended to improve the mothers' ability to obtain and keep a job, including adult education and employability development classes), and (4) total program participation.

Even after controlling for baseline characteristics that *predicted* each of these aspects of program participation, certain parenting measures continued to be significantly associated with the extent of program participation among mothers in the experimental group. Thus, within this group, scores on the HOME-SF Emotional Support subscale continued to differ significantly according to extent of participation as defined in each of the four ways; and the observational measure of Book Reading Quality differed according to participation in the human capital development components of the program as well as in the program overall.

It should be noted that characteristics we did not have baseline measures of, and thus could not control for, might be linked both to greater program participation and to parenting behavior, and these undocumented variables might help account for the associations we have noted. These findings nevertheless raise the possibility that components of programs for young mothers in poverty *including but going beyond* parenting education may have positive implications for parenting behavior. Further study is needed to explore how and why program components directed at mothers' education or employment skills, and their overall program participation, might affect their parenting behavior.

IX. Parenting Behavior as a Predictor of Child Outcomes

The findings on child outcomes in the New Chance Demonstration present us with a paradox. Although there were positive impacts on measures of parenting, children in the experimental group did *not* do better in terms of their cognitive and social development when these were assessed at the 42-month follow-up. On most measures, there were no program impacts. However, on some measures there were unanticipated *negative* program impacts. For example, mothers in the experimental group described their children as having less positive social behavior. This pattern held in both the full demonstration study sample and the smaller observational study sample.

Clearly, we need to consider factors other than parenting behavior as contributors to children's development — for example, mothers' psychological well-being, family economic status, and the experiences of mothers and children outside the mother-child relationship. New Chance appears to have had unexpected negative program impacts on some of these further factors.[4] For example, mothers in the experimental group in the observational study sample were more, rather than less, depressed at the time of the first (18-month) follow-up survey. Mothers in the experimental group of the observational study sample also reported less life satisfaction.

We considered both the role of parenting behavior and variables that measured other important aspects of the mothers' and children's lives as predictors of five selected child outcomes within the observational study sample (children's total scores on the School Readiness Component of the Bracken Basic Concept Scale, the Behavior Problems Index total score as reported by both mother and teacher, and the Positive Behavior Index total score as reported by both mother and teacher). We confirmed that when measures of parenting behavior were taken into account, our ability to predict children's scores on several of these outcomes improved significantly (specifically, the maternal report of the Behavior Problems Index and Positive Behavior Index and the teacher report of the Positive Behavior Index).

---

[4] See Janet C. Quint et al., *New Chance: Interim Findings on a Comprehensive Program for Disadvantaged Young Mothers and Their Children* (New York: MDRC, 1994), Chapter 6, and Quint, Bos, and Polit, 1997, Chapter 6.

More supportive and stimulating parenting thus predicted more optimal developmental outcomes on specific measures of development. However, in addition, we found that variables reflecting maternal psychological well-being (for example, measures of life satisfaction and of aggravation and stress in parenting) and variables reflecting the larger social context of the families (for example, measures of difficult life circumstances and of number of changes of residence since enrolling in the evaluation) were also significant predictors of several child outcomes. Greater maternal psychological distress and greater stress in the larger social context predicted less positive developmental outcomes.

An important finding of this study, then, is that while parenting behavior was a significant predictor of specific child outcomes, it was not the only predictor. The positive influence of supportive and stimulating parenting behavior *combined* with negative influences of psychological distress in the mother and stress in the family's larger social environment. Children's developmental outcomes reflect influences not only from within but also outside the mother-child dyad.

Outcomes for children in the context of an intervention such as New Chance might improve if the intervention directly addressed these problems through more intensive mental health intervention for the mothers; if the "dosage" of program components with positive implications for children's development, such as parenting education, were substantially increased; or if program elements targeted to the children themselves were strengthened. Direct observations of New Chance child care settings placed these child care centers just below the "good" range in terms of quality.[5] For positive child outcomes to occur, children from high-risk families may need higher-quality care. In addition, beyond the first program follow-up (at 18 months), children in the New Chance experimental group were not found to participate in more child care than those in the control group. Indeed, the increased child care participation of children in the experimental group tended to occur only during the first months of their mothers' program participation, when the mothers were engaged in on-site classes and activities. Positive program impacts on children may require *sustained* participation in high-quality child care.

X. How Observational Measures of Parenting Contribute to a Study Such as the Evaluation of New Chance

Measures of parenting based on direct observation of mother-child interaction contribute to our understanding of the New Chance Program in several ways. First, these measures *increase our certainty about program impacts on parenting*. If positive impacts on parenting were found only for maternal report measures, we might question them on the grounds that mothers in the experimental group (aware that they had access to New Chance Program services and that outcomes of the program were being assessed) could report more favorably because they felt it to be expected of them. Even interviewers aware of which research group a family was in could be subject to such "response biases." Coders of the videotapes were unaware of which research group a family was in, yet positive program impacts were found on observational measures.

---

[5] Quint et al., 1994, pp. 73-76.

Second, observational measures *provided information that was different from and complementary* to that provided by interview-based measures of parenting. For example, a positive program impact was found on the observational measure of Book Reading Quality, yet no program impacts were found on the interview-based measure of cognitive stimulation (the HOME-SF Cognitive Stimulation subscale). The observational measure looks at the *nature or quality* of mother-child interaction in a book reading context, for example, the mother's fluency, intonation, and comfort level in reading the book to her child. By contrast, the HOME-SF measure of cognitive stimulation assesses *the quantity of literacy-related and other stimulating activities:* for example, how *many* books were in the home, and how *often* the mother reads to the child. Without a measure of the nature or quality of cognitive stimulation, we might not have known that there were program impacts in this area. Correlations between observational and interview-based measures of parenting confirm that these measures are related but do not substantially overlap.

Finally, *we found the observational measures to be helpful in predicting child outcomes.* When we distinguished among maternal report scales, HOME-SF subscales, and observational measures as predictors of the selected child outcomes, we found that *the HOME subscales and observational measures were generally better predictors than the maternal report scales.* We also asked whether observational measures *added* to our ability to predict variation in children's developmental outcomes after the maternal report scales and HOME-SF subscales were already taken into account. We found that for two of the five selected child outcomes, *observational measures added significantly to our ability to predict child outcomes even when the other parenting measures had been taken into consideration.*

The fact that the observational measures do not rely at all on maternal report also helps eliminate the possibility that associations between parenting behavior and those child outcomes based on maternal report (for example, the Behavior Problems Index, as reported by the mother) are not merely a reflection of common attitudinal or reporting tendencies across different sets of maternal report measures.

Thus, observational measures of parenting behavior provide different information than that available through interview-based measures of parenting, diminish concerns about possible response biases and correlated measurement error, and add significantly to the ability to predict specific child outcomes. Where the examination of parenting behavior is a high priority, observational measures of parenting add substantially to the strength of an evaluation.

Part II: Methodological Assessment of the New Chance Observational Study

I. The Methodological Context for the New Chance Observational Study

The New Chance Observational Study lies at the confluence of rising interest in policy-relevant research among developmental psychologists, interventions focusing on two generations in a family, and increasing demand for nontraditional forms of survey research. In attempting to increase the size and representativeness of samples, some developmentalists are departing from the familiar model of laboratory-based research, sometimes supplemented with home visits, to undertake their data collection through contracted survey research. Survey research on parenting and child development has traditionally involved querying parents about their parenting practices and their children's development. In recent years, survey interviewers have also been asked to rate parent-child interactions and the home environment, and, in a few cases, to administer structured parent-child interactions similar to those usually carried out in the child development laboratory. New Chance, while not alone in its use of survey methods to study child development, is one of the few studies to use survey interviewers to conduct observational work with mothers and children. Within this context, Part II of the monograph seeks to familiarize readers with the "survey model," document how the observational study was conducted, assess the success of the effort, and consider both specific recommendations for future work and broader implications for research design.

This study affords a unique opportunity for examining methodological issues in the measurement of parenting and child outcomes for two reasons. First, it taps a diverse set of measures and data sources: taped mother-child interactions coded under rigorous conditions in university laboratories, self-reports of mothers, and ratings by survey interviewers of mother-child interactions and the home environment. Just as important as the rich variety of domains tapped is the fact that the study was conducted within the framework of a survey research model, while "stretching" this model to exploit it in innovative ways.

II. Implications of Conducting Research Within a Survey Model

Chapter 10 of the monograph seeks to enhance the value and accessibility of the survey field to social scientists who have not made survey research the primary focus of their careers and to begin to bridge the communication gap between survey researchers and their colleagues — even from the same disciplines — who may emphasize a substantive research agenda more than a method. Without a full appreciation of the constraints of the survey model, researchers who commission survey work may have difficulty managing it and even find themselves disappointed with the results. At the same time, they will be shortchanged if they look to the field primarily for a data collection capability and fail to heed its lessons on survey-based measurement.

In contrast to an academically oriented model, survey research is built on a division of labor between those responsible for the conceptual and analytical aspects of research and those who actually collect the data. The survey model places responsibility for data collection in the hands of "distant proxies" for "absent researchers." These proxies, the survey interviewers who collect data on surveys ranging from small local studies to large recurring government surveys, are neither assumed to possess nor expected to master a conceptual appreciation of the research in which they participate. For these reasons, survey researchers provide interviewers with explicit rules to direct them through the data collection — rules that, in theory at least, require minimal judgment to apply. The resulting survey model is characterized by an orientation toward production, a contrived and stylized format for the interaction of interviewer and respondent, and precise programming and standardization of interviewer behaviors. Designing sound data collection instruments and procedures involves accommodating to these three aspects of the survey model.

III. Turning to the Survey Field for Help with Measurement Issues

The constraints, or "rules," of the survey model may be viewed as potential limitations in conducting social science research, but the survey community's trove of research on survey measurement can be viewed as a valuable resource. The underlying focus of most of this work is the reduction of error in surveys.

The bulk of the measurement literature has focused on the design of measurement instruments, that is, questionnaires, and less on other aspects of measurement, such as interviewer or mode effects. If this literature yields an overarching lesson, it is a humbling one: that there are many, many features of questions and questionnaires that can affect survey response. An awareness of such measurement threats is helpful both in developing measures and related data collection instruments and in assessing their likely strengths and limitations.

The survey literature also speaks to the ways in which interviewers may affect measurement in a study like this. Evidence for the prominence of interviewer effects in situations requiring more judgment or unprogrammed behavior is especially relevant to the present work, which required interviewers to make substantive ratings of the home environment and mother-child interaction, obtain time-use information through a series of open-ended probes, and administer a scripted observational protocol, while applying general principles to unscripted situations.

Given the importance of interviewers and the expansion of their responsibilities with ever more challenging studies, focusing on the cognitive demands of their tasks will help in assessing what is sensible to ask them to do and helping them to do it. The last decade or so has seen the emergence of a broader interest in cognitive aspects of survey response, beyond simply the recall of information. This interest has been directed primarily toward the tasks facing *respondents* during an interview, such as the strategies they use to estimate and report the frequency of a particular behavior. The cognitive framework can also be applied to the cognitive tasks faced by *interviewers* during a survey interaction, by examining, for example, the ability of interviewers to process the information necessary for making substantive ratings at the same time they are reading interview questions and recording answers. Viewing the demands of the survey interaction on both respondents and interviewers from a cognitive perspective provides a helpful framework for integrating what is known about the instruments used to collect data, the interviewers who administer them, and the respondents who provide the data.

IV. Steps in Implementing the New Chance Observational Study

As discussed in Chapter 11, and summarized below, implementing the observational research that is the central subject of this monograph involved both the design of data collection instruments and procedures and the data collection effort itself. In carrying out these tasks, we confronted and sought to address a host of survey measurement issues, most notably the overarching issue of adherence to the goals and objectives of the "absent researchers" who place their research in the hands of survey interviewers. Our approach included the following steps.

- **Tasks developed in university laboratories were adapted for survey administration by strictly scripting them in the format of a structured survey questionnaire to ensure standardized delivery by survey interviewers.**

Survey interviewers are not expected to have the appreciation of a study's theoretical underpinnings that would allow them to work from only a researcher's semistructured outline. Therefore, the strategy for realizing the objectives of the "absent researchers" was to program interviewer behavior in the observational session carefully. This meant providing interviewers with a data collection instrument resembling a survey questionnaire, which specified the precise language to be used and the actions to be taken. The instrument included detailed instructions for presenting, arranging, and withdrawing the various props used in the tasks (book, games, and gift) and for coordinating management of the props with the script. Also included were language and instructions for verifying that the mothers understood the objectives of the task, decision rules for determining how much time to spend on each, and, for one task, decision rules as to whether a simple or complex version, or both versions, of the task was to be administered.

- **The potentially unlimited number of situations that could not be scripted explicitly were addressed by developing a set of general principles that survey interviewers were expected to apply.**

A survey questionnaire typically tries to anticipate and provide for every scenario that can be tapped by a question; this is generally done by creating response categories for recording answers to a question that are exhaustive and (usually) mutually exclusive and by employing "skip patterns," in which questions are asked or not depending on the answers to prior questions or characteristics of the respondent. Like response categories, skip patterns must be designed to anticipate all of the situations to be encountered.

By contrast, the data collection instrument for the observational study could not anticipate the virtually infinite range of behaviors and events that might occur during the session, not only involving the mother and child but also other people and events in the household. Since it would have been impractical to embed instructions for responding to even a sampling of hypothetical situations in the data collection instrument, a separate set of guiding principles, rules, and examples was developed, which interviewers were expected to apply in the actual data collection situation. Although our intent was to minimize the need for judgment on the part of the interviewers, they did have to label a situation correctly in order to respond appropriately, which increased the complexity of the task. For instance, they needed to understand the distinction between a mother's request for clarification of the instructions or objectives of a task and a request for guidance as to how to work with her child, since these two situations called for quite different responses.

- **Although the activities that constituted the observational session had enjoyed extensive use in laboratory-based research, pretesting was necessary to investigate issues specific to survey administration.**

Two small pretests were conducted to help in refining the adaptation of the laboratory protocols for survey administration, check on the effectiveness of the tasks among young disadvantaged respondents, and investigate a number of practical implementation issues. Even these small pretests anticipated a variety of situations and issues that would arise again in the main study. For instance, the home environments in which the data collection occurred were typically characterized by a host of distractions and interruptions from television and radio, children, and other sources. The pretests also revealed how the interviewers' personal styles could affect the way in which the sessions were administered. For instance, some interviewers seemed to have particular difficulty maintaining a neutral, nondirective stance. Making sure the mother understood each task first arose as a concern during the pretesting and remained an issue during the data collection, since interviewers were charged with the somewhat incompatible goals of clarifying the objectives of the task if the mother failed to understand it, without telling the mother how she should work with her child. Requirements that interviewers make subtle distinctions — reinforce effort rather than performance, for example — posed problems during the pretests and continued to do so during the study. Many other issues were successfully resolved through pretesting, however, and rarely resurfaced.

- **Training materials and curricula, developed through the design and pretesting process and supplemented by ongoing quality control, provided the critical link between the researchers designing the work and the lay survey interviewers collecting the data.**

Training materials. The basic resources included a procedures manual for the study and a training videotape, which supplemented training that interviewers had already received for the New Chance 18-month survey to make the HOME-SF (Home Observation for Measurement of the Environment-Short Form) and related ratings also called for by the observational study.

The procedures manual developed for the study was similar in many ways to manuals used for more typical surveys, covering a mix of substantive and administrative requirements and emphasizing use of the data collection instrument. The manual also addressed topics specific to the observational study such as the ways in which the assignment resembled and differed from traditional survey interviewing, special requirements for conducting the observational session, principles for relating to the child during the visit, definitions of questionnaire items not covered in training materials for the 18-month survey, detailed guidelines for handling unscripted situations, coordination in the field between the two members of the data collection team (interviewer and videographer), and technical instructions related to the taping. The manual was designed not only for training the data collection teams but also to serve as a reference resource throughout the study.

Training curriculum. The first component of interviewer training consisted of completion of a home-study package. Trainees were expected to study the procedures manual and review a sample videotape. In addition, interviewers were expected to practice delivering the observational script and complete a quiz covering the major requirements of the study.

A one-day "classroom" training session followed. A joint session of interviewers and videographers began with an overview of the project, followed by a critical viewing of the training videotape, with discussion of the points it illustrated. In the next module, interviewers and videographers practiced working together in assembling and removing the various props. The joint session concluded with a discussion by the group of the need and mechanisms for close communication between interviewers and videographers.

In a separate session later in the day, videographers were introduced to the specific equipment to be used in the study and received training on their technical responsibilities. A simultaneous session for interviewers began with a review of the maternal self-report items used in the observational study questionnaire but not in the 18-month survey, which were to be completed during the interview segment of the observational session. This was followed by a review of the quizzes completed as part of home study. Then, with trainers playing the parts of mothers and children, each interviewer administered two "mock" observational sessions that had been scripted in advance to include a variety of situations for them to handle.

After the classroom session and as the last step in the training, each interviewer completed a "real" practice observational session with a family in the community similar to those in the New Chance sample. Interviewers and videographers received feedback after review of their work by members of the observational study team before beginning work with actual sample members.

V. Findings on the Administration of the Observational Sessions

In reflecting on this work, the challenges that faced the data collection teams and the performance of survey interviewers in a nontraditional role were of particular interest. Interviewers were to administer the protocol in such a way that all the mothers began with a clear idea of the objectives of each task and then maintain a polite professionalism and neutral stance while each dyad worked together. The challenges to mothers and children of working together on tasks that required some "stretch" for the children were not to be magnified by problems created by the interviewer, such as failing to explain a task adequately or creating additional anxiety by emphasizing the child's performance. Nor were the inherent challenges of the tasks to be *reduced* by "helpful" suggestions from the interviewer. Other extraneous influences that we had hoped to minimize were intrusions by other people or events during the observational session.

Chapter 12 of the monograph presents findings on how the observational sessions were actually administered. These findings draw on a variety of data sources: interviewer ratings, maternal self-reports, comments about the session and its administration recorded by the university coders, and variables coded in the university laboratories.

- **The ideal environment for conducting the sessions was rarely encountered. The sessions were typically conducted with persons besides the mother and child present, and interruptions and distractions were commonplace.**

While interviewers tried to schedule the sessions when only the mother and the focal child were at home, they were generally unable to do so: in 70.9 percent of the cases, there was someone else present. Children other than the focal child were most common, being present in 51.0 percent of the cases. Although the presence of others did not guarantee interruptions of the session, this was often the effect. In 33.8 percent of the sessions another child or children interrupted the session at least once; in 7.6 percent of the sessions there were three or more such interruptions. In addition, background noise was often present: in 46.6 percent of the cases there was audible background conversation, while a radio, television, or stereo contributed background noise in 29.3 percent of the cases.

- **A proxy measure of interviewer performance, representing overall competence in administering the observational session, suggests that a large majority of the sessions were administered satisfactorily, if not optimally.**

Comments from the Harvard University laboratory on the two tasks coded there — the book reading and discussion activity and the wheels task — served as a proxy for overall interviewer performance. Based on these comments, we conclude that there was certainly room for improvement in the interviewers' performance: in 42.0 percent of the sessions, the coders noted some deviation from the protocol. However, in about half of these cases, only a relatively minor deviation on one of the two tasks was noted. Thus, overall, 79.9 percent of the sessions were characterized by either no deviations or only a relatively minor deviation on one of the two tasks. Still, a deviation judged to be at least moderately serious was noted in about 10 percent of administrations of each task. Many of the deviations cited by the coders clearly represent errors by the interviewers, in that they disregarded instructions or principles contained in the various study materials and addressed in interviewer training. The most common of these, seen in 8.9 percent of the sessions for the book activity and 7.2 percent for the wheels task, involved offering direction, intervening in the dyad's work, or otherwise abandoning a neutral stance in some way. Other problems arose either because interviewers were expected to make difficult judgments in the press of the situation or because they were discouraged from exercising their own judgment, sometimes resulting in behavior that appeared awkward or inappropriate when viewed on tape.

- **Considerable variation was observed in the proficiency with which the survey interviewers administered the sessions, with the more proficient conducting more sessions.**

The number of sessions completed by each interviewer ranged from 1 to 52. There was a positive association between interviewer proficiency and the volume of cases completed, with more proficient interviewers completing more cases. This finding may have resulted from retraining and practice effects or the selection out of the less proficient (or less committed) interviewers.

- **Despite concern about the potential intrusiveness of a study like this, coupled with the mild stress created by the challenge of the tasks, most mothers found the session to be a positive experience.**

Examination of the mothers' self-reported experience of the observational session revealed that 54.1 percent rated it in the very positive range on a composite measure, with virtually all of the rest rating it in the intermediate range. But when focusing only on mothers' perceptions of the more challenging performance-related aspects of the experience (for example, nervousness, difficulty working with her child), 37.9 percent reported a very positive experience and 14.6 percent expressed quite negative feelings.

- **The age of the focal child was the central mediator of the experience of the session for both mother and child and was related to the frequency of interviewer errors as well.**

While the various indicators of the experience of the session that were examined were associated with one another, the child's age was the common thread that united them, with the experience being more negative the younger the child. The interviewer's difficulty in conducting the session also apparently related to the age of the child, with interviewer errors being more frequent the younger the child.

VI. Methodological Implications of the New Chance Observational Study

We have asked whether it is possible to carry out observational work within a contracted survey research model and, in Chapter 13 of the monograph, answer the question with a qualified "yes." We argue for an approach built on explicit recognition of the differences between the backgrounds of survey interviewers and developmental psychology researchers, as well as recognition of the cognitive demands of this work. The cognitive processes employed by interviewers in responding to unprogrammed situations, in which they must exercise judgment because the situations cannot be precisely scripted, are of particular interest. To optimize the value of the data collected through survey methods, it is necessary to recognize the complexity of the interviewer's assignment and to pay careful attention both to the design and testing of data collection protocols and to the selection, training, and supervision of interviewers. Careful design efforts should be followed by a formal interviewer certification process, with ongoing quality control and feedback to interviewers on their performance.

But interviewer effects represent an ever present source of potential bias despite efforts undertaken to minimize them. This is true for all measures with which they are associated, whether derived from interviews, interviewer ratings, or administration of an observational protocol. To *measure* the extent of such effects, random assignment of (fairly large numbers of) cases to interviewers is necessary, but difficult and costly to accomplish. However, it is possible to *control* for interviewer effects in a study employing an experimental design, by ensuring that each interviewer's caseload mirrors the distribution of experimentals and controls in the overall sample.

VII. Implications for Measurement Strategies in Studying Parenting and Child Outcomes

Researchers faced with choices about the measures and methods to employ in studying parenting and child outcomes require a framework for making decisions about study design. As discussed in Chapter 14 of the monograph, such a framework can encompass both methodological and substantive considerations. *Methodological* considerations include feasibility and data quality. Based on the New Chance Observational Study, as well as other work, we conclude that developmentalists can draw on a methodological repertoire within a survey framework that includes maternal self-reports on parenting, interviewer ratings, cognitive assessments of children, interviews with children, and administration of observational protocols. The availability of such a broad range of measurement techniques means that researchers can assemble a "balanced portfolio" of measures that spread the measurement "risk" across methods having different strengths and vulnerabilities.

The overarching *substantive* consideration in asking whether to include observational work within this portfolio is the "value added" of the observational data. Such work could be undertaken as part of a freestanding study, in which all participants provide both self-reported data, perhaps enhanced by interviewer observations, and observational data. Or observational work could be *embedded* within a larger survey effort, as was the case in this study. In considering the embedded model, the "analytical leverage" provided by the observational data is a consideration. If the subsample participating in the observational component of the study is selected randomly, it is possible to generalize to the larger sample, and/or to the population from which it was drawn, through statistical weighting procedures. Other potential analytical strategies include using the richness of the observational measures to help in interpreting data from other sources, including assessing how much confidence to place in other measures that are potentially subject to different method effects. Finally, observational measures from a subsample may be used to augment quantitative survey findings.

Researchers, however, should not expect to find a single decision rule for determining when observational research within a survey model is warranted. Every study is different. Survey-based observational work makes sense when it serves well-articulated analytical objectives and is conceptualized as an integral part of an overall research design. If resources are available and the potential added value of observational measures is clear for a particular study, the observational method is a valuable adjunct to more conventional survey approaches.

Part I: The New Chance Observational Study

Chapter 1

Introduction: The Context for the New Chance Observational Study

Martha J. Zaslow

The New Chance Observational Study provides detailed and sensitive measures of mother-child interaction for a subset of the families who participated in the full evaluation of the New Chance Program. This observational study addressed four broad issues: (1) whether the New Chance Program, a comprehensive and supportive intervention for teenage welfare mothers with limited education, had positive impacts on parenting behavior; (2) how measures of parenting collected by direct observation and through interviews are related, and how these two kinds of parenting measures differentially predict child outcomes; (3) what role parenting measures play, relative to such other important factors as mothers' psychological well-being and family economic self-sufficiency, in shaping the development of the young children of these mothers; and (4) what methodological issues researchers need to be aware of when they study parenting and child outcomes through a survey research approach that draws on interview-based measures (maternal self-report and interviewer ratings) as well as measures based on direct observation of mother-child interactions.

This first chapter of the monograph provides the context for the New Chance Observational Study by describing the evaluation within which this special study is embedded. We summarize the goals and components of the New Chance Intervention, as well as the results available from the full New Chance Evaluation. The chapter concludes by noting the specific questions that will be addressed in the observational study monograph and by providing an overview of its organization.

I. Goals of the New Chance Observational Study

This monograph presents the results of the New Chance Observational Study, a special study embedded within the evaluation of the New Chance Intervention.[1] Four broad issues were addressed by this study:

---

[1] Throughout this monograph we use the terms "New Chance Program" or "New Chance Intervention" to refer to the classes, counseling, and further supportive services offered to New Chance participants and their children. The "New Chance Evaluation" includes the assessment of program implementation, cost-benefit analysis, and assessment of both short- and longer-term program impacts on mothers and children. The "New Chance Demonstration" encompasses both the program and its evaluation. The New Chance Intervention and Evaluation will be described in detail later in the chapter.

- **whether a comprehensive intervention for young mothers who are high school dropouts receiving welfare improves their parenting practices, as measured through direct observation of mother-child interaction and through interview-based measures of parenting;**
- **how closely measures of parenting behavior collected via direct observation of mother-child interaction and through interviews with mothers are related, and how well each kind of parenting measure predicts child outcomes;**
- **what role parenting behavior plays, relative to other important influences (such as maternal psychological well-being and the family's economic self-sufficiency), in shaping the development of the young children of mothers participating in this intervention; and**
- **what methodological issues are raised when researchers augment a program evaluation by embedding observational research in a survey research context.**

In this observational study, 290 of the 2,322 mothers in the New Chance Evaluation received an extra visit to their homes to collect data following the 18-month follow-up within the larger evaluation. The extra visit, carried out on average 21 months[2] after enrollment in the New Chance Demonstration, involved videotaping mothers interacting with one of their children while reading a book, carrying out a series of teaching tasks, and presenting a small gift to the child. The children videotaped interacting with their mothers were between 27 and 63 months old with most between 30 and 60 months. Two sets of measures were derived from coding of the videotapes: measures of the affective quality of mother-child interaction, and measures of mother-child interaction related to the development of literacy. A brief interview with the mothers at the time of the observational session provided measures of mothers' allocation of time to parenting and further data on such key issues as use of child care and current employment. The interview carried out 18 months after enrollment for the full evaluation sample (including the observational study sample) provided further interview-based measures of parenting, derived from maternal self-report interview items as well as interviewer ratings of the home environment.

The New Chance Evaluation provides a unique context for asking *whether* a comprehensive program for young welfare mothers can affect parenting behavior, *how important* parenting behavior is relative to other factors in shaping child outcomes, and *how* to go about measuring parenting behavior within such an evaluation. Several characteristics of the New Chance Program distinguish it from other efforts to improve the life circumstances of mothers on welfare. First, the intervention itself, in several important ways, went beyond previous interventions for families in poverty. For example, New Chance focused on a population that, while of critical importance for welfare policy, is little studied: that subgroup of young mothers on welfare who have already dropped out of school. This group is at risk in terms of *both* its parenting behavior (with implications for the well-being of the next generation) and its long-term economic well-being. Second, New Chance was one of only a handful of new interventions for families in poverty that followed a "two-generational" approach (Smith, 1995b): that is, an approach that seeks to support the development of both parent and child, rather than focusing on one generation alone.

---

[2] The plan for the study called for the in-home data collection effort of the observational study to be carried out as soon as possible following completion of the 18-month follow-up, and within four months of that follow-up, within the larger evaluation. While the observational session occurred 21 months after random assignment, on average, it ranged from 16 to 36 months after random assignment.

Perhaps most important, while other interventions focus on enhancing mothers' parenting behavior *or* their economic circumstances, New Chance sought to enhance both aspects of mothers' lives simultaneously. The program included parenting classes and incorporated material regarding parenting behavior in several further program components, such as life skills training classes, family planning classes, and counseling. The program also included multiple program components aimed at strengthening the young mothers' employability. Mothers started with adult basic education classes, but went on to employment training classes, apprenticeships, and job placement activities.

As one of very few interventions that sought to facilitate both the parenting and employment roles simultaneously, the New Chance Intervention provokes a series of important questions. The young mothers who participated in New Chance needed to proceed with lives in which they were both parents *and* sources of economic support to their families. On the one hand, perhaps young mothers thrive in a program environment that explicitly recognizes the balancing act they need to master in their lives and, accordingly, make progress in both spheres. On the other hand, in a population that is at high risk by dint of young age, poverty, and difficult life circumstances, perhaps it is overly ambitious to seek to enhance both roles simultaneously. Does the need to incorporate classes and workshops addressing both the maternal and employment roles, as well as program components for the children, into the intervention's programming dilute the program's dosage to an extent that undermines each set of priorities? Were these at-risk mothers stressed rather than aided by the comprehensiveness of the program and its expectation of change in multiple dimensions of their lives?

Within this intervention and for this population, then, it is particularly important that we ask whether parenting behavior improved and whether other data for the same families indicate that it improved along with positive changes in mothers' education, employment skills, and employment. Given the richness of the data available on a range of further possible influences on children's development in this sample, it is also particularly important that we ask about the role that parenting behavior played in shaping child outcomes relative to such other factors as children's experiences in child care, mothers' psychological well-being, mothers' welfare receipt, educational attainment, acquisition of job skills, employment, and earnings.

In addition to the substantive issues that can be addressed by embedding observational research within the larger New Chance Evaluation, doing so also creates an unusual opportunity for examining the *process* of measuring parenting behavior and the *measures* themselves. Most observational studies of parent-child interaction have been carried out with small samples (samples of 30 to 50 families are not unusual in such research; see, for example, Denham, Renwick, and Holt, 1991; Landry et al., 1990). Yet such observational research may be an important addition to the set of measures in multisite evaluation research with a focus on parenting. This approach may help to diminish both problems that occur when predictors (parenting measures) and outcomes (measures of children's development) come from the same source (maternal report),[3] and response biases in parenting measures that request information directly from mothers or interviewers. This approach may also yield more detailed and qualitative measures of the mother-child relationship.

At the same time, carrying out observational research within the framework of a larger evaluation provides an opportunity to examine observational measures of parenting in light of extremely rich data from interviews on the mothers and families. It is rare for observational studies to include information with the amount of detail available in the present study on such issues as the families' economic circumstances, residential situation, mothers' education attainment and literacy, and mothers' psychological well-being. Because the follow-up surveys of the New Chance Evaluation provide interview-based measures of parenting, it is also possible to look at the relationship of observation- and interview-based parenting measures.

There are a variety of ways to implement observational research in a multisite study. One way is to develop a collaboration among university research groups and *rely upon these research groups themselves to collect the data*. In such a strategy, professors and their students at the collaborating university sites participate in the development of a common observational study protocol and agree upon standardized procedures for collecting and coding the data. Each university team implements the protocol within a sample at its site. The videotapes from all sites are then coded jointly. This strategy was adopted by the National Institute of Child Health and Human Development Study of Early Child Care (Friedman, 1995). A second way to extend the use of observational procedures is to *rely upon survey interviewers for the collection of observational study data*. In this strategy, research teams again develop the observational study protocol. However, they also adapt the protocol so that it can be administered by survey interviewers. The observational study data are then collected by survey interviewers as part of their other data collection activities within a sample. This strategy was adopted by the study of parent-child interactions embedded within the Teenage Parent Demonstration (Aber, Goodman, and Morris, 1993; Brooks-Gunn and Berlin, 1993).

The New Chance Observational Study chose the second strategy, but with a commitment from the start to examine the feasibility, problems, and advantages of this strategy. In order to carry out this self-reflection, measures have been included in the embedded study to examine the issue of mothers' subjective responses to the completion of the observational procedures in this manner. In addition, following completion of the study, we summarized available records on the extent to which the survey interviewers adhered to and departed from the carefully scripted observational study procedures (see Chapter 12 of this monograph).

This monograph is organized around the four broad issues that the New Chance Observational Study sought to address. Part I provides a detailed examination of program impacts on measures of the mother-child relationship (issue 1), contrasts measures of parenting derived from observational and interview contexts (issue 2), and examines the implications of parenting behavior, relative to other factors, for child outcomes (issue 3). Part II provides an assessment of the strategy of implementing an observational study using a survey research approach (issue 4).

---

[3] This problem, correlated error associated with consistent method effects, is discussed in some detail in Chapter 10.

In order to fully understand the goals, procedures, and findings of the New Chance Observational Study, it is necessary first to understand the context of this embedded study: the population served by the New Chance Program, the intervention strategy itself, the timing and content of the full evaluation, and the findings available from the larger evaluation. We include in this overview a summary of findings not only from the first follow-up of the New Chance Evaluation, but also recently released findings from the final follow-up of the evaluation. Some unexpected findings from this final follow-up, particularly concerning children's development, underscore the importance of understanding how the New Chance Program affected parenting behavior and how parenting behavior, in turn, contributed to developmental outcomes.

We turn now to an overview of the New Chance Program and Evaluation, including a summary of earlier and recently released findings (see Quint et al., 1994; and Quint, Bos, and Polit, 1997). In concluding the introduction, we use the information from this overview to specify, in greater detail, the questions that are addressed in the two parts of this monograph.

II. The New Chance Intervention

New Chance targeted mothers between ages 16 and 22 who were receiving public assistance, who gave birth at age 19 or younger, and who had dropped out of high school. For this target population, New Chance provided a comprehensive and supportive program that aimed to help these mothers complete the General Educational Development (GED) certificate and go on to develop employment skills and enter employment. In addition to enhancing the mothers' economic self-sufficiency, New Chance also supported the mothers' personal development and health through life skills education, family planning and health education classes, personal and group counseling, and facilitating their access to health care services. As we have noted, New Chance was a two-generation program that sought to enhance the development of both the young mothers and their children. While mothers were active in the program, their children were entitled to child care services. In addition, the program sought to improve access to health care services for the children as well as for the mothers. Of particular importance in the present context, mothers in the New Chance Program participated in parenting education classes.

The New Chance Evaluation, like the intervention itself, focused on two generations rather than one. The evaluation considered program impacts on children as well as mothers. It also entailed separate consideration of shorter- and longer-term program effects, with follow-up interviews carried out both 18 and 42 months after mothers enrolled.

A. Special Characteristics of the New Chance Target Population

The target population of the New Chance Program is of particular importance from a policy perspective. In recent years, adolescent mothers have constituted only about 8 percent of the overall welfare caseload in this country (Granger, 1994; 1995).[4] Yet giving birth as an unwed teenager places a mother at high risk of eventual poverty and welfare receipt. Examinations of the welfare population in the past decade have revealed that a substantial proportion were teenagers *at the time of the birth of their first child*. For example, Moore (1990) found that 59 percent of women receiving Aid to Families with Dependent Children (AFDC) had been teenagers when their first child was born. Estimates in the past decade indicate that approximately three out of four adolescent mothers received welfare within five years of the birth of their first child (Adams, 1990). Further, nonwhite women who had not completed high school and who began receiving welfare as unmarried mothers were particularly likely to receive welfare benefits for a sustained period. Bane and Ellwood (1983) found that the average welfare duration for this group was 10 years. Finally, the population of unwed teenage mothers is growing. The number of unmarried women who had a birth between ages 15 and 19 was 87,100 in 1960. By 1992 this number had increased to 353,878 (Brown and Stagner, 1996).

Thus, it is of national importance to consider effective means for enhancing the long-term economic self-sufficiency of those who became unwed mothers as teenagers. Policymakers repeatedly note the importance of this target population. Nevertheless, young mothers who are high school dropouts have not in fact been widely served. Few states have focused attention on serving adolescent mothers with young children. Further, those that do serve them primarily target those who are still in school (Quint et al., 1994).

The distinction between adolescent mothers who are still in school and those who have dropped out is important (Granger, 1994; 1995). In a review of the evidence from programs serving poor adolescent mothers, Granger concluded that impacts differ for "students" (those who have graduated from high school or are still in school) and for those who have dropped out before completing high school. Demonstration programs report some positive impacts for adolescent mothers who are students, but give little cause for optimism among dropouts. For example, Granger (1995) reports that positive five-year impacts on employment and earnings among the adolescent mothers participating in Project Redirection were driven entirely by the subset that was still in school at the onset of the program.

The fact that few intervention efforts have been addressed to a group of such policy importance is attributable at least in part to the difficulty and complexity of serving unwed teenage mothers who have dropped out of school (Granger, 1994; 1995). Quint and colleagues (Quint et al., 1994; Quint and Egeland, 1995) point out that the child care needs, limited education, and limited history of employment in the group of young high school dropouts make them a group with poor prospects of employment as well as costly to serve. Further, data from the New Chance Evaluation make it clear that these young mothers face difficult life circumstances that make such a target population unusually hard to serve.

New Chance staff documented extremely high rates of serious problems among the young mothers in the program (Quint and Egeland, 1995). Nearly half (48 percent) had such serious housing problems as being evicted, being asked by the primary tenant to leave, or lacking a permanent residence entirely. Seventeen percent had family members who were involved in substance abuse; 12 percent of the mothers themselves used alcohol in a way that interfered with program participation and 15 percent used illegal drugs in such a manner. Physical and emotional abuse touched the lives of many of the young mothers: 12 percent had experienced physical abuse as children and 10 percent sexual abuse. Sixteen percent were reported or observed to be battered by a current partner and 6 percent by someone other than the current partner. The young women were sometimes discouraged by key individuals in their lives from participating in the program: 15 percent by partners and 9 percent by their own mothers.

---

[4] The nation's largest cash welfare program, Aid to Families with Dependent Children (AFDC), was replaced by the Temporary Assistance for Needy Families (TANF) program under legislation signed into law in August 1996. The characteristics of welfare recipients under TANF are not yet known.

> Strikingly, almost three-quarters had experienced at least one of these unquestionably serious situations.... The personal problems faced by the large majority of New Chance clients makes them an indisputably hard-to-serve group (Quint and Egeland, 1995, p. 107).

While dysfunctional parenting in the family of origin was not documented in a quantitative manner by New Chance staff, a qualitative study of 50 mothers who had participated in New Chance (34 of whom had completed their GEDs and 16 of whom had not) identified this as a key issue. Discussing the group that left New Chance without completing the GED in this descriptive study, Quint and Musick (1994) note that

> the majority of these young women, while they were growing up and throughout their adolescent years, have suffered continuously poor nurturing, repeated separations (or outright abandonment), or loss of mothers or other emotionally significant parental figures. Only five of the 16 young mothers were raised consistently by their own mothers.

The unmet needs for nurturance by some of the young New Chance mothers served to undermine their steps toward psychological and economic autonomy and their ability to parent their own children.

In sum, New Chance targeted a critical population, but one burdened by serious problems (including for some a lack of supportive parenting in the family of origin), and one for whom there has been limited effective programming. The descriptions of parenting in the families of origin of some New Chance participants highlight both the importance of addressing parenting behavior within this intervention and the possible difficulty of doing so effectively.

We turn now to a description of the New Chance Program *model*: how the program was intended to unfold, the nature of program services, and the program's short- and long-term goals. In the larger evaluation, Quint and colleagues have distinguished between the program model and the actual experience of program participants — for example, documenting discrepancies between intended and actual levels of participation (Quint, Fink, and Rowser, 1991; Quint et al., 1994; Quint, Bos, and Polit, 1997.). In our summary of findings from the larger New Chance Evaluation below, we briefly summarize these findings on actual participation for the full evaluation sample. In addition, in Chapter 3 we document actual levels of participation in New Chance Program components for families in the embedded observational study, focusing especially on those aspects of the program hypothesized to affect parenting. Here, however, our focus is on the program "blueprint."

B. Program Phases

In its goal of seeking to improve the employability of the young mothers in New Chance, this program looked beyond the completion of the GED to consider the eventual employment of program participants. As a result, the program had two phases. Phase 1, lasting several months, emphasized completion of the GED. Mothers were expected to attend 20-30 hours per week of classes, with mornings (2-3 hours) devoted to adult basic education and GED preparation and afternoons to a variety of other classes and workshops (including parenting classes; see below). Phase 2 was initiated when mothers had completed the GED or been in the program for five months. The goal of this phase was to help mothers enter and keep jobs with benefits and the possibility of advancement, through a choice of vocational training, work internships, and further education. Throughout both phases, a case manager worked with each mother, monitoring activities and, especially in phase 2, guiding the choice of activities and contacts with other agencies. The program provided for up to 18 months of program participation, and an additional year of follow-up case management.

C. Program Sites

New Chance was piloted at 6 sites and actually implemented at 16 sites in 10 states. The program was successfully implemented by organizations with various service traditions, including community service organizations, government agencies, adult high schools, and community colleges (Quint, Fink, and Rowser, 1991).

D. Program Services

Program guidelines specified types and amounts of services to be delivered at each site. Program services fell into three broad categories: (1) human capital development, (2) personal development of the mothers, and (3) development of the children.

1. Human Capital Development Services. Services aimed at mothers' human capital development, or skills and knowledge needed for employment, were a primary focus of the New Chance Program and were allocated a high proportion of program time and resources. During phase 1, mothers participated in classes aimed at successful completion of the GED. Mothers also participated in classes aimed at employment and career awareness, which focused on a variety of occupations and the skills needed to enter them. In phase 2, mothers participated in occupational skills training tailored especially to jobs in the local market, full or part-time work internships, job placement activities, and further education at community colleges.

2. Personal Development Services. Personal development classes and workshops aimed to address the many problems that the young women faced in their everyday lives. Life skills training focused on mothers' communication and decision-making skills in sexual contacts, relationships, parenting situations, and work. Health education classes and health services included components on nutrition, stress management, physical and emotional abuse, substance abuse, childhood immunization, sexually transmitted diseases, and depression and anxiety. Family planning issues, addressed in both group sessions and individual counseling, aimed at encouraging the young women to delay subsequent pregnancy and childbirth. Mothers also received individual counseling through meetings with their case managers. The young mothers met individually with their case managers once every two weeks during phase 1 of the program and generally less often during phase 2. Together, clients and case managers developed an individual program plan and reassessed progress periodically. Program guidelines recommended caseloads of no more than 25 clients (15 clients, if case managers had other responsibilities).

3. Child Development Services. Children were provided access to pediatric health services on an as-needed basis, mostly at off-site clinics and hospitals. New Chance offered child care at no cost to all participants for as long as they remained active in the program. In 12 of the 16 program sites such care was offered on site, although for two of these, only drop-in care was available. Sites that arranged for off-site care for clients' children also monitored the quality of care provided. Guidelines for on-site and off-site child care called for programs that were supportive to young children's development.

Of particular importance in the present context, mothers in New Chance participated in parenting education classes, which involved a balance of (1) presentations aimed at covering specific content and (2) group discussions directed by the concerns of participants rather than a specific curriculum. Program guidelines called for coverage of such issues as stages of development (physical, emotional, and cognitive), developmentally appropriate cognitive stimulation, strategies for setting reasonable behavioral limits at different ages, and building self-esteem and trust in children. Guidelines also called for joint mother-child sessions once a month to provide a context in which staff could demonstrate activities and interactions, observe mothers interacting with their children, and provide feedback.

E. Short-Term and Longer-Term Program Goals

New Chance had both short-term and longer-term program goals. In the short term, the program components summarized above aimed to enhance maternal educational attainment, health status, emotional well-being, parenting behavior, and the mother-child relationship; also, to help mothers postpone further pregnancy or childbearing and to expose children to high-quality child care. Longer-term goals included increasing mothers' employment and earnings, reducing welfare receipt, and improving children's development across the domains of health, cognitive development, and social development.

III. The New Chance Evaluation

A. The Evaluation Design

A rigorous evaluation of the New Chance Program has been conducted by the Manpower Demonstration Research Corporation (MDRC), consisting of three key components: an implementation evaluation, an impact evaluation (that is, an examination of differences between experimental and control group families on outcome variables), and a cost-benefit evaluation. It is crucial to note that the impact analysis, unlike many other program evaluations focusing on low-income families, encompasses measures of the well-being of both the mothers and their young children. Information was collected from the mothers just prior to their assignment to an experimental or control group (at "baseline") and through in-home interviews at 18 and 42 months after enrollment in the evaluation. Findings from the implementation study (Quint, Fink, and Rowser, 1991) and the 18-month survey (Quint et al., 1994) have already been published. Findings from the 42-month survey, including data regarding children's cognitive development, socioemotional development, and health, have just been released (Quint, Bos, and Polit, 1997).

Each of the program sites was asked to recruit approximately 150 participants for the evaluation. Those mothers who volunteered for the program, met the eligibility criteria,[5] and agreed to participate in the evaluation were randomly assigned to an experimental or control group, with approximately two-thirds of those enrolled in the evaluation being assigned to the experimental group and one-third to the control group. (A two-to-one random assignment ratio was adopted because it was reasoned that sites would respond more favorably to random assignment, and probably would be able to recruit applicants more readily if young women knew that their chance of admission to the program was better than 50-50. This ratio also facilitated the examination of findings for key subgroups within the experimental group.)

The evaluation sample was enrolled in the study between August 1989 and July 1991. Altogether 2,322 mothers were enrolled. When a mother had more than one child, one was randomly chosen prior to the 18-month follow-up as the "focal child," or child to be focused upon in consideration of program impacts on children. Experimental group mothers were enrolled in the New Chance Program. Control group mothers were not allowed to participate in New Chance, but were free to seek other services in the community. Impact analyses in the New Chance Evaluation thus consider not how New Chance affected mothers and children compared with families receiving *no* relevant services, but rather how New Chance affected families above and beyond the services that the control group families sought and received on their own.

Of the 2,322 mothers enrolled in the evaluation sample, 2,106, or 90.7 percent, were interviewed in their homes approximately 18 months after enrollment (the first follow-up). For the second follow-up, carried out approximately 42 months after enrollment, 2,105, or 90.6 percent, of the mothers initially enrolled were interviewed in their homes. These response rates are extremely high for a sample of high-risk families.[6]

In addition to the measures designed to capture program impacts for mothers, the final follow-up of the New Chance Evaluation provides a number of measures of the children's development. These outcomes, measured approximately 42 months after mothers enrolled in the demonstration as part of the final follow-up, permit consideration of program impacts on the children.[7] During the 42-month interview, a direct assessment of the children's cognitive development, the School Readiness Component of the Bracken Basic Concept Scale (Bracken, 1984) was administered. This scale is a measure of receptive language with content reflecting school readiness. The subtests that constitute the School Readiness Component of the scale assess the specific concept areas of colors, letter identification, numbers/counting, comparisons, and shapes. For the New Chance Evaluation sample, internal consistency reliability for this measure is quite high at .98. As an indication of the validity of the measure, for those children in the New Chance Evaluation sample whose teachers completed the Teacher Questionnaire, the total score on the School Readiness Component varied significantly by teacher rating of children's academic performance (Polit, 1996b).

---

[5] Eligibility for the program was determined through an initial interview, during which data on the young women's characteristics at baseline were recorded on the New Chance Enrollment Form. A reading test (the reading part of the Tests of Adult Basic Education, TABE) was also administered to program applicants at baseline (see Quint et al., 1994, p. 25, for further details).

[6] Data were sufficiently complete to permit analyses for 2,088 of the 2,106 interviews completed at the 18-month follow-up and for 2,079 of the 2,105 interviews completed at the 42-month follow-up.

[7] See Polit (1996a) for a detailed description of each measure, reasons for selection of the measure within the New Chance Evaluation, and reliability and validity data.

To assess child outcomes in the area of sociemotional development, two measures were obtained from mothers: the Behavior Problems Index (Zill, 1985) and the Positive Behavior Index (Polit, 1996a). These two measures were also obtained through a survey of the teachers for those focal children in the sample already attending school, Head Start, or a structured, academically oriented preschool program. Teacher survey data are available for 835 or 79.5 percent of these children (86.2 percent of the 969 children whose mothers allowed their children's teachers be contacted). The Behavior Problems Index asks an adult who is familiar with a particular child to report on the incidence of problems in six areas: antisocial, anxious/depressed, headstrong, hyperactive, dependent, and peer conflict/withdrawn behavior. The index yields a total score as well as subscale scores for each of the six behavior areas. This measure was chosen for the New Chance Evaluation because it is widely used (for example, in the Child Health Supplement of the National Health Interview Survey and the Child Supplement of the National Longitudinal Survey of Youth) and has been shown to have strong reliability and validity in previous work. Within the New Chance Evaluation sample, children whose mothers reported a discipline problem that resulted in a communication from school had higher average total Behavior Problems Index scores. Scores on the index were also related to interviewer ratings of the children's behavior during the 42-month follow-up (Polit, 1996a).

The Positive Behavior Index was developed specifically for the New Chance Evaluation when a review of existing measures found none that both addressed positive social behavior in children and could be considered appropriate for administration in a disadvantaged sample (Polit, 1996a). The Positive Behavior Index provides a total score and scores for three subscales: compliance, social competence, and autonomy. Internal consistency reliability coefficients for this newly developed measure ranged from .77 to .94 for maternal report. The total score on the Positive Behavior Index was again related to the report of a communication from school regarding a discipline problem and to the interviewer's ratings of the child's behavior.

In addition to the Behavior Problems Index and the Positive Behavior Index, mothers and teachers provided a series of further ratings. Mothers rated children's academic performance, overall health and use of health services. Teachers provided ratings on children's general academic performance and ability, self-esteem, motivation, overall school adjustment, and how well they got along with teachers and students. Teachers also reported on disciplinary action taken and whether they had had meetings with mothers for a general discussion of their children. Many of the items from the New Chance Teacher Questionnaire were taken or adapted from the Teacher Questionnaire used as part of the 1981 National Survey of Children (Polit, 1996b).

B. Key Findings from the Implementation Study

MDRC's report on the implementation of New Chance (Quint, Fink, and Rowser, 1991) notes that despite their varying service traditions, the 16 New Chance sites successfully mounted all phase 1 activities. The sites found education among the easiest of the phase 1 services to put in place, and employability development and individualized family planning services the most difficult. Phase 2 services, because they were offered off site and through a variety of outside services, were harder to implement in a uniform manner. The outside agencies did not have the same familiarity and skill in working with the population targeted by New Chance. Further, New Chance case managers found it difficult to continue their biweekly contacts with their off-site clients given their on-site caseloads.

Analyses indicate discrepancies between intended and actual levels of program participation. Absenteeism was a problem in the New Chance Program, with approximately half of those enrolled actually attending program activities on a typical day. While the young mothers in the experimental group clearly received more services than those in the control group, control group mothers sought and received non-New Chance services in the community on their own initiative; for example, as reported in the New Chance 18-month follow-up, while 85 percent of experimental group mothers participated in some education classes, 60 percent of control group mothers did as well. The discrepancy in participation was greater for program components *other* than education; for example, approximately three times as many mothers in the experimental group as in the control group participated in parenting education classes. These findings are in keeping with the fact that mothers who volunteered for New Chance were primarily seeking education rather than personal development services.

Table 1.1 illustrates the discrepancy between intended and actual program participation. This table compares mean hours of actual program participation with hours of program participation specified in program guidelines for experimental group mothers over the first 18 months of program participation. As can be seen, while program guidelines called for between 32 and 64 hours of participation in parenting classes, on average this sample of mothers participated for less than 18 hours. The discrepancy between intended and actual hours of program participation does not reflect a problem with service *delivery;* all sites provided services to the extent specified. Rather, the discrepancy reflects attendance issues. Perhaps in keeping with the many difficulties in their lives, and the voluntary, rather than mandatory, nature of the program, New Chance mothers showed inconsistent attendance.

Differences in levels of program participation were, not unexpectedly, linked to background characteristics of the mothers. Experimental group mothers who participated more in New Chance Program activities had more positive educational histories (higher educational attainment, less likelihood of having repeated a grade), were more likely to have parents who had graduated from high school or completed a GED, were more likely to be using birth control and to have received medical care on program entry, and reported more emotional support. Among racial/ethnic subgroups in the evaluation, Hispanic mothers attended New Chance activities most, followed by black mothers and then white mothers.

Analyses of program costs at 18 months indicate that on average the program cost the

Table 1.1

Program Guidelines for Participation and Mean Hours of Actual Participation in New Chance Within 18-Month Follow-Up Period

Program Component	Number of Hours Specified by Program Guidelines	Mean Hours of Participation in Component
Education	192	101.0
Employability development	48	26.7
Family planning	6	6.3
Health education	24	10.5
Parenting education	32-64	17.6
Life skills	36-39	20.5
Other group activities (e.g., field trips or social events)	---	19.3
Average hours of participation in all components[a]	---	297.6
Sample size		1,408

Sources: New Chance Management Information System (MIS) data. Program guidelines from MDRC 1991 implementation report.

Notes: Calculations for this table used data for 1,408 experimentals for whom there were 18-month follow-up survey data, including values of zero for those who were randomly assigned to New Chance but did not participate.

Dashes indicate unavailable data.

[a] Excludes individual counseling and college classes.

sponsoring agencies $5,073 per mother in the experimental group, excluding the cost of child care. Child care cost an additional $2,573 on average, and costs from other agencies totaled an average of $1,380.

C. Characteristics of the Evaluation Sample at Baseline

Examination of the characteristics of the evaluation sample[8] at the time they enrolled in the evaluation confirms that this was indeed a group of disadvantaged young mothers with limited education. This group of mothers averaged just under 19 years of age at the time they enrolled in the demonstration. On average, they had given birth to their first child before age 17. More than three-fourths of the sample (77 percent) were from minority racial/ethnic groups, and fewer than 1 in 10 had ever been married. About one-third of the mothers already had two children or more, and a majority had a child under age 2. A little over one-third of the mothers had dropped out of school before their first pregnancy. Mothers averaged more than two years out of school and had reading skills just above the 8th grade level. Few mothers had worked full time for a sustained period. Nearly two-thirds of them (64 percent) had grown up in a family with some welfare receipt. Ninety-five percent of the mothers were currently receiving AFDC, nearly all as heads of their own households. A much higher proportion of mothers in this sample than in broad community-based samples (53 percent), received scores on a depression index indicating that they were at risk for clinical depression.

D. Key Findings from the Impact Evaluation: The 18- and 42-Month Follow-Ups[9]

1. Education and Literacy. Analyses of 18-month program impacts — that is, differences between experimental and control group mothers across particular outcomes at this time — indicate that some of the articulated short-term program goals but not others were effectively addressed. A higher proportion of mothers in the experimental group than in the control group (37 percent as opposed to 21 percent) had completed the GED, and more mothers in the experimental group had some college credits by the time of the 18-month follow-up. Yet no group differences were found in mothers' literacy levels, with about 40 percent of each group reading at or below the 7th grade level. Mothers averaged a 7.6 grade reading level.

At the 42-month follow-up, several markers of educational attainment continued to point to positive program impacts. The proportion of mothers who had completed the GED grew for both the experimental and control groups between the 18- and 42-month follow-ups, but the proportion of mothers with a GED continued to be significantly higher in the experimental group (45 percent versus 33 percent). A significantly higher proportion of mothers in the experimental group had either a GED or a high school diploma (52 percent versus 44 percent) and, further, a higher proportion of mothers in the experimental group had completed some college credits (14 percent versus 11 percent).[10] At the same time, a smaller proportion of mothers in the New Chance experimental group (7 percent) than in the control group (10 percent) had obtained a high school diploma, perhaps because of the program's emphasis on completing the GED.

---

[8] The figures for baseline characteristics of the sample in this paragraph are based on data for the 2,088 respondents included in analyses of the 18-month survey as reported in Quint et al., 1994. For baseline characteristics of the full New Chance sample of 2,322, see Quint et al., 1994, Appendix B. In that report, statistical tests showed no large or systematic differences between the two samples.

[9] In the following section, 18-month findings are from the New Chance interim report (Quint et al., 1994). Forty-two-month findings are from the final report (Quint, Bos, and Polit, 1997). The final report also includes selected 18-month outcomes analyzed for the sample of families included in the 42-month report. As this was a slightly different sample from that used in the analyses for the interim report, the numbers reported here from the 18-month interim report differ slightly from the selected 18-month findings presented in the 42-month report.

2. Markers of Economic Self-Sufficiency. Findings on economic self-sufficiency at the 18-month follow-up for the New Chance sample suggest an "opportunity cost" of participating in educational activities. That is, in the short run, participation appeared to be associated with lower rates of employment and lower earnings, as mothers initially focused on education rather than employment. At 18 months, *control group* mothers had higher cumulative earnings. More control group mothers had been employed in the first months after random assignment. Across the 18-month follow-up period, 43 percent of mothers in the experimental group and 45 percent of those in the control group had worked at some point. Over 80 percent of mothers in both groups were still on welfare at 18 months. Experimental group mothers were more likely to be living in public housing or receiving rent assistance.

Although earnings and labor market experience are forgone during periods of investment in education and job skills training, the expectation is that this initial cost will eventually yield higher rates of employment and better earnings. Yet at the time of the 42-month follow-up, no group differences were found for mothers' employment or earnings. Experimental and control groups did not differ significantly in terms of the proportion of mothers employed at all or employed full time. Further, no significant differences were found in earnings or hourly wages. Without a further follow-up for the New Chance Intervention planned, it is impossible to say whether this pattern will hold or whether the employment patterns will diverge again in the future, with mothers in the experimental group eventually benefiting from their increased educational attainment.

3. Fertility. At the time of the 18-month follow-up, there was no indication that experimental group mothers had successfully delayed subsequent childbearing. More than one quarter of mothers in both groups had had a subsequent child at this follow-up point. Although there were no group differences in terms of childbirth at the time of the 18-month follow-up, experimental group mothers actually had had a significantly higher rate of pregnancy and abortion than control group mothers.

By the time of the 42-month follow-up, there were no indications of important differences in the fertility behaviors of mothers in the experimental and control groups. The proportion of mothers who had had a pregnancy or birth did not differ significantly by group, nor did the average number of post-enrollment pregnancies. At the 42-month follow-up, mothers in the experimental and control groups reported similar patterns of sexual activity and contraceptive use (for example, the proportion of mothers reporting that they were sexually abstinent and not pregnant, or sexually active and contracepting regularly, did not differ across groups). There is no evidence, then, that the New Chance Program had impacts on mothers' fertility behaviors.

---

[10] This difference, while numerically small, was statistically significant.

4. **Residence.** During the first 18 months of the evaluation, mothers in the experimental group showed more shifts in residence than mothers in the control group. At the time of the 18-month follow-up, mothers in the experimental group were more likely to be living with a partner or husband and less likely to be living with a parent or grandparent. Residence and fertility patterns were linked. A higher rate of subsequent pregnancy among experimentals occurred only for those living with a husband or partner.

Findings from the 42-month follow-up provide further indications of a group difference in residence patterns. The average number of times that mothers had moved (from the time of the child's birth to the time of the 42-month follow-up) was significantly higher in the experimental group (averages of 4.0 and 3.7 times in the experimental and control groups, respectively). Further, more mothers in the experimental group reported having had trouble finding a good place to live in the past year (41.9 and 37.5 percent in the experimental and control groups, respectively), and more experimental group mothers were living without any of their children at the 42-month follow-up (4.9 percent as opposed to 2.8 percent). At the second follow-up there were no longer significant differences in the proportion of mothers living with a husband or partner or living with a parent or grandparent. Yet a small but significantly higher proportion of mothers in the experimental group were living in "other" arrangements at 42 months (living with friends, alone, or in an institution).

5. **Maternal Psychological Well-Being.** Findings regarding maternal psychological well-being differed substantially for the first and second follow-up points. At 18 months, no program impacts were detected regarding maternal depression, stress, sense of control over events, or health. However, mothers in the two groups differed in their reports on two aspects of social support. Experimental group mothers were less likely than control group mothers to say that they had no one to turn to for emotional support; these mothers also indicated that they were more satisfied with the social support that they were receiving.

By 42 months, however, there were indications that New Chance had had negative impacts on mothers' psychological well-being. Average scores on the measure of depression were significantly higher in the experimental group at this later time. While depression scores declined in both the experimental and control groups over time from baseline to 42 months, the decline was significantly greater for the control group mothers. A higher proportion of mothers in the experimental group reported that they felt stressed much or all of the time in the past month, and their satisfaction with standard of living was significantly lower. At the same time, there were no indications at the final follow-up point that mothers in the experimental group felt a greater sense of social support.

Several possible bases of these unexpected findings on maternal subjective well-being are discussed in detail in the full report of 42-month findings (Quint, Bos, and Polit, 1997). One possibility is that New Chance raised expectations for improvements in life circumstances, but realizing these improvements proved much more difficult than expected. As a result, New Chance mothers may have experienced "dashed hopes," manifested in greater stress, depression, and dissatisfaction with economic circumstances.

6. Parenting Behavior As Documented in the Context of Interviews. At 18 months, statistically significant differences were found on two of the interview-based measures of parenting. On a measure of control/punitiveness, experimental group mothers reported less harsh and more authoritative (that is, supportive but firm) disciplinary attitudes. Further, on the component of the Home Observation for Measurement of the Environment (HOME) Inventory (short form) that measures mothers' provision of emotional support to the focal child, experimental group mothers were found to be providing more emotional support.

At 42 months, no overall group differences were found on the subscales of the HOME Inventory. A positive program impact on the HOME total score was found, however, for the subgroup of mothers who did not show indications of depression when they entered the program. In keeping with the greater indications of stress and depression among experimental group mothers, mothers' report of stress in parenting was higher in the experimental group at 42 months (as measured by both a total parenting stress score and a subscale of aggravation with child rearing).

7. Child Care. By the 18-month follow-up, children in the experimental group were substantially more likely to have experienced nonmaternal care, especially center-based child care. Experimental group children were also more likely to have been in a regular child care arrangement before age 1. This pattern reflects, at least in part, the fact that many New Chance sites offered on-site center day care.

By the 42-month follow-up, the two groups had converged in terms of reliance on child care. No group differences were found in terms of current use of a child care arrangement, or participation specifically in center child care, family day care, regular care by a grandparent, Head Start, or school. While a higher proportion of children in the experimental group had participated in a day care center or preschool over the full course of the evaluation, this difference was largely attributable to the earlier months of program participation.

8. Child Outcomes. Child outcomes are available from the 42-month follow-up and the teacher survey. Again unexpectedly, negative program impacts were detected across several measures. Mothers' reports of children's behavior problems were significantly higher in the experimental group (overall and on three of six Behavior Problems Index subscales). At the same time, mothers' reports of children's positive behaviors were significantly lower in the experimental group (overall and for all three Positive Behavior Index subscales). There were no group differences in mothers' overall ratings of their children's health, in the proportion of children who had been hospitalized at least once, or in the proportion of children who had had an emotional trauma. Yet mothers in the experimental group reported that a higher proportion of their children had had an injury, poisoning, or accident requiring medical attention.

As we have noted, for some of the children in the evaluation sample, ratings of behavior problems and strengths were available for the same children from both the mother and a teacher. Within this smaller sample of children already attending a school or academically oriented preschool, significant differences between experimental and control groups emerged more often in maternal reports regarding the child than in teacher reports. Specifically, mothers in the experimental group rated the academic performance of their children less favorably than mothers in the control group. In addition, more mothers in the experimental group reported being notified by the school of behavior problems. By contrast, teachers rated the academic performance of the children in the two groups similarly. Teachers rated *daughters* of mothers in the experimental group as presenting more behavior management problems requiring parental notification, but saw no group difference in the behavior of *sons*. The teacher survey found no group differences on a series of measures that only the teachers reported: teachers saw no significant difference in the proportion of children in the experimental and control groups who had problems getting along with students or teachers, with self-esteem or motivation, or in terms of their overall adjustment.

The fact that indications of negative program impacts on children's behavior were seen more through the eyes of the mothers than the teachers deserves further scrutiny. Several interpretations are possible for this discrepancy in child outcome findings by source. One possible explanation focuses on *context*. Children manifest different behaviors at home and at school. Perhaps differences in the behavior of children in the experimental and control groups surface more often in the context of the home environment. Such an interpretation seems plausible particularly if children are more restrained in the context of the greater structure and behavioral expectations of school, or if problem behaviors that are reported more often in the experimental group are directed specifically toward the mother or other family members.

A second interpretation focuses on the *informant*: mothers' *perceptions* of their children's behaviors may have been affected by participation in New Chance, whether or not the children's behavior actually differed. As we noted earlier, New Chance mothers exhibited higher levels of depression than control group mothers at 42 months, which might have colored their perceptions of their children. Another possibility concerns day care participation. Because more children of mothers in the New Chance experimental group than in the control group had been in day care, mothers in the experimental group had had, over time, more contact with child care providers. It is possible that conversations with child care providers alerted mothers to problems regarding their children's social behavior that child care providers thought needed to be addressed within a classroom context. Such problems might have been occurring with equal frequency among children who had not attended day care (and with equal frequency among children of mothers in the experimental and control groups). Yet with more frequent contact with child care providers, mothers in the experimental group may have been alerted to such problems with greater frequency. When asked about their children's behavior, they may have therefore been more likely to acknowledge the presence of behavior problems. Thus, participation in New Chance might have altered the perception of behavior if not actual child behavior.

A third interpretation focuses on *sample*. The sample for the teacher survey differed from the full evaluation sample in that children were somewhat older. In addition, all of the children in the teacher survey sample were currently participating in some form of academically oriented school or preschool. Children in academically oriented settings and their families may have different characteristics from those not making use of such programs (Zaslow et al., 1997). Perhaps the characteristics that predispose some families to the use of educational settings also buffer against negative program impacts on child outcomes as they would be manifested in school settings and reported on by teachers.

E. Summary of Findings

These complex and sobering findings from the full New Chance Evaluation speak to the difficulty of working with the group of very young welfare mothers who have dropped out of school. Despite the many obstacles facing this group of mothers, the New Chance Program was able to bring about positive changes in educational attainment. Yet 42 months after the start of the program there was no indication that the enhanced educational attainment had resulted in an improvement in mothers' economic self-sufficiency. The program was also unable to alter mothers' fertility behavior. Further, there are unexpected findings that New Chance had negative implications for mothers' psychological well-being and residential stability.

In the present context, the findings on parenting behavior and child outcomes from the larger study are particularly salient and they present a paradox. On the one hand, there are modest indications that parenting behavior was positively affected by the New Chance Program. Positive impacts were detected at the first follow-up on two interview-based measures of the emotional quality of the mother-child relationship. Yet despite these indications of changes for the better in parenting behavior, there were negative program impacts on several important child outcomes. Mothers of children in the experimental group perceived their children as having more behavior problems and as showing fewer positive social behaviors. Mothers of school-age children in the experimental group described their children as doing less well in school both academically and behaviorally.

A substantial body of research documents that parenting behavior is of great importance in shaping child outcomes (reviewed in Maccoby and Martin, 1983; Bornstein, 1995). Why would there be negative impacts on child outcomes when New Chance seems to have contributed to a positive impact, albeit small, on parenting behavior as measured by interview-based measures of parenting? Several explanations seem plausible. First, perhaps the positive program impacts that the program had on parenting at the first follow-up were of too small a magnitude to have influenced subsequent child outcomes in a substantial and positive manner. Second, perhaps other aspects of the families' lives (such as maternal stress and depression or economic deprivation) were important in shaping child outcomes and overwhelmed any positive implications that the program may have had for parenting. Third, perhaps we should scrutinize and question the parenting measures collected as part of the two follow-up interviews. Measures of parenting that rely on the reports of mothers and ratings of interviewers may be subject to a variety of response biases. For example, the fact that these informants were aware of the assignment of families to the experimental group versus the control group may have affected the ratings they gave. Perhaps a markedly different picture of program impacts would emerge (such as an absence of group differences or a negative program impact on parenting) if measures of parenting behavior were based on the ratings of outside observers, especially those unaware of whether families had or had not participated in New Chance.

The embedded observational study makes it possible to weigh these various possibilities for a subset of the full evaluation sample: that program impacts on parenting were not of sufficient magnitude to contribute positively to child outcomes, that other factors overwhelmed a positive influence of parenting on child outcomes, and that more rigorous parenting measures would yield a different picture of program impacts on parenting. The observational study provides ratings of both the affective and cognitive aspects of mother-child interaction, completed by observers less vulnerable to a number of possible response biases, under conditions requiring intensive training in the constructs and measures of interest, and high interrater reliability. We are thus able in this monograph to extend the examination of program impacts on parenting behavior to a broader set of parenting variables, documented in previous work to be highly sensitive to family context and predictive of child outcomes, and with fewer concerns about response bias. In particular, we ask in subsequent chapters whether the pattern of modest program impacts on parenting is corroborated with the observational measures and whether the highly sensitive observational measures tell us anything more about the particular kinds of behaviors affected by the program. With access to measures of parenting derived from both interview and observational settings, the current monograph also examines in detail the issue of whether and how parenting behavior measured in different ways contributes to child outcomes. Finally, we access the broad set of variables available from the two follow-up interviews in the full evaluation to ask how other factors combine with parenting variables in shaping the child outcomes.

IV. Questions Addressed by the New Chance Observational Study

A. Questions Concerning Program Impacts on Parenting Behavior and the Role of Parenting Behavior in Shaping Child Outcomes

In order to examine the implications of the New Chance Program for parenting behavior, and the role of parenting behavior in turn in shaping child outcomes, the monograph addresses the following sequence of questions:

- **How can we describe the parenting behavior of the mothers in the New Chance Observational Study overall?** Using methods and procedures that have been employed in other studies of at-risk populations, what can be said broadly about the affective quality and stimulation for early literacy in mother-child interactions?

- **Which subgroups appear to be at greatest risk in terms of their parenting behavior, and which are providing the most stimulating and supportive interactions?** Although the mothers in the New Chance Observational Study are uniformly young, poor, and of limited education, is there nevertheless meaningful variation within this sample in terms of parenting behavior? Do the affective quality and literacy-stimulating aspects of interaction vary according to background characteristics such as whether or not the mother received welfare as a child, the age at which the mother gave birth to her first child, and the years of schooling she has completed? If there are groups that appear particularly strong or particularly at risk in terms of parenting behavior, are the patterns the same for the literacy-related and affective dimension of interaction?

- **Did the New Chance Intervention have positive impacts on the observational measures of parenting behavior?** If so, did it have positive impacts across the two domains of parenting behavior examined through direct observation of mother-child interaction — the affective quality of interactions and the aspects of interaction related to the development of literacy — or were impacts restricted to one of these key domains? Within each domain, are effects detected across multiple parent and child behaviors, or are effects more circumscribed? Are findings for the observational measures concordant with the program impact results for the interview-based measures of parenting within the observational study sample?

- **If positive impacts are detected, did they occur consistently across families with varying characteristics, or were they found primarily for definable subgroups?** For example, are positive impacts documented for those mothers who could be said to be at less risk at the start of the evaluation (for example, mothers with only one child, with more years of education and less indication of school failure) or also for those at greater risk?

- **How do positive impacts come about?** If New Chance has positive impacts on parenting behavior, can we trace such impacts to the parenting education component of the intervention, or are other program components also contributing? Are there different explanatory processes for different types of parenting measures, or do all of the measures follow the same pattern?

- **How do the observational and interview-based measures of parenting relate?** What is the degree of overlap among these measures? What distinctive information, if any, is provided by the observational measures? When we add observational measures of parenting to an evaluation, do we improve our ability to predict child outcomes?

- **How important are measures of parenting, relative to other sets of variables, in predicting the cognitive and social development of the young children in the sample?** How do such further variables as maternal educational attainment and employment status, maternal depression, current residence pattern, and child care history combine with the parenting variables to influence child outcomes?

B. Questions Concerning the Methodology Used to Document Parenting Behavior

We have noted that the New Chance Observational Study followed a relatively new fielding strategy in order to introduce the intensive procedures of observational research into a multisite evaluation. The present study reflects the intersection of two traditions: the attempts of developmental psychology to extend observational research to larger and more representative samples and of survey research to go beyond interview formats to other data collection strategies. Our methodology can be seen as a "hybrid" approach that seeks to use the strengths of the survey model in tracking and maintaining contact with a population that is difficult to locate and include in research, and the strengths of developmental psychology in developing tasks and coding procedures for recording interactive behaviors that are of central importance to children's development.

The New Chance Observational Study not only attempted this new approach, but also incorporated tools to evaluate the effectiveness of the strategy. We have mothers' reactions to the observational session (for example, ratings of such items as "the interviewer helped me feel comfortable during the session" and "while I was working with my child on the book and games, I felt nervous about how I was doing"), interviewers' reports of the context for the observational session (for example, noise level, interruptions from other children), and coders' comments about the extent to which the script and procedures for the observational sessions were adhered to or violated during each session.

In the second part of this monograph, we critically examine our methodology of carrying out observational procedures within a survey research framework, asking specifically:

- **What are the implications of conducting observational research through a survey research model?** As developmentalists turn to contracted survey research in order to leverage larger and/or more representative samples, they will need to understand the culture of survey research and how this may differ from laboratory-based research in order to derive the greatest value from the collaboration.

- **How can the body of methodological work on survey measurement inform the study of parenting behavior using a combination of interviews, interviewer reports, and observational procedures?** While turning to the survey community for its data collection capabilities and enhanced samples, developmentalists may also wish to draw on the survey field's corpus of methodological work on survey measurement. While most of this work focuses on respondents' answering questions from interviewers, there are important implications for observational research as well.

- **How were the laboratory-based observational protocols adapted for survey administration, and how was the observational study implemented in the field? What challenges were encountered in conducting the observational study in this way?** The lay survey interviewers who collected the data lacked substantive background and familiarity with previous research using the observational protocols. They needed a highly structured set of data collection materials, including a "script" resembling a survey questionnaire, in order to present a standardized stimulus situation. This work presented logistical challenges in carrying out the field work (for example, coordinating the work of two-person teams), situational problems (for example, interruptions from other children needing attention, lack of work space), and management difficulties associated with the poor "fit" of this type of study with ongoing survey operations.

- **How well did lay survey interviewers meet the challenges of administering observational research?** The observational study required survey interviewers to extend their repertoires significantly. What have we learned about the ability of interviewers to handle such a task and about their potential for such work in the future, when the lessons from the current effort can be incorporated?

- **How did the mothers experience the observational session? What were the mediators of participants' subjective experience of the session?** The tasks were designed to be sufficiently challenging for the children to need their mothers' help, but not to be unpleasant or overly stressful. How successful were we at creating this balance? A number of variables could contribute to mothers' sense of stress in the situation or to having had a positive experience, such as the interviewer's skill at administering the protocol, the interviewer's experience in interviewing disadvantaged young women, and the age of the child. Which, if any, of these are important?
- **What recommendations can we make for future research involving obtaining observational measures of parenting within a survey research context?**

V. The Organization of the Monograph

Part I of this monograph examines the first set of questions concerning program impacts on parenting behavior, the relation between observation- and interview-based measures of parenting, and the role of parenting behavior in shaping child outcomes. Chapter 2 describes the sample and procedures used in the New Chance Observational Study. Chapter 3 presents data for the observational study sample regarding program participation, focusing especially on those program components with some content relevant to parenting. The chapter also provides qualitative data from retrospective interviews with field staff regarding the content of the parenting education component. Chapters 4 and 5 present findings on program impacts on the affective quality of mother-child interaction and literacy-related aspects of observed interaction, respectively. Chapter 6 completes the picture of program impacts on parenting measures by summarizing findings for the observational study sample on the interview-based measures of parenting. Chapter 7 looks across all of the available measures of parenting, comparing the magnitude of program impacts on each of the parenting measures; considering which program elements, if any, the program impacts on parenting variables were linked to; asking how closely the differing parenting measures are correlated; and asking whether the observational measures add significantly to our ability to predict child outcomes beyond the measures of parenting derived from the interview context. Chapter 8 asks what role parenting behaviors — along with variables reflecting mothers' human capital, mothers' psychological well-being, and families' current context — play in shaping child outcomes. Chapter 9 concludes Part I with a summary of findings and a discussion of their implications.

Part II focuses on the measurement process. Chapter 10 explores the implications of conducting research on parenting, particularly observational research, using a survey research approach, and considers how methodological research on survey measurement can inform such work. Chapter 11 discusses the process of translating protocols from developmental psychology into a format suitable for survey administration and summarizes the challenges that were encountered in fielding the study in this manner. Chapter 12 presents data on the quality of the survey interviewers' work and mothers' subjective reactions to the observational session. Chapters 13 and 14 conclude Part II by providing an evaluation of the success of our fielding strategy and extracting key recommendations for future research seeking to bridge the distance between intensive observational procedures and larger, more representative samples.

Throughout the monograph we draw heavily on previous MDRC publications regarding the New Chance Program and Evaluation. We rely especially on *New Chance: Implementing a Comprehensive Program for Disadvantaged Young Mothers and Their Children* (Quint, Fink, and Rowser, 1991); *New Chance: Interim Findings on a Comprehensive Program for Disadvantaged Young Mothers and Their Children* (Quint et al., 1994); *Lives of Promise, Lives of Pain: Young Mothers After New Chance* (Quint and Musick, 1994), and *New Chance: Final Report on a Comprehensive Program for Young Mothers in Poverty and Their Children* (Quint, Bos, and Polit, 1997). We also draw heavily on *Two-Generation Programs for Families in Poverty: A New Intervention Strategy* (Smith, 1995), which includes a chapter by Quint and Egeland on New Chance.

Chapter 2

The Methodology of the New Chance Observational Study

Donna R. Morrison, Carolyn A. Eldred, Martha J. Zaslow, and M. Robin Dion

This chapter presents the methodology of the New Chance Observational Study. We begin by describing the sample for the observational study: how the 290 families who participated in this study were selected, and what their characteristics were at baseline. Next we turn to an examination of program impacts on variables from the 18- and 42-month follow-up surveys. Whereas Chapter 1 provided a summary of program impacts from these surveys for the full New Chance Evaluation sample, here we present impact findings specifically for the observational study sample regarding maternal education, employment and economic self-sufficiency, fertility, and psychological well-being, as well as child outcomes. We then provide an overview of the procedures of the observational study, detailing the mother-child interaction tasks and the brief interview that took place at the time of the observational session, approximately 21 months after families enrolled in the New Chance Evaluation. Finally, the chapter summarizes all of the parenting measures available for the observational study sample and gives an overview of the strategy of analysis used to address the specific questions of the New Chance Observational Study.

Chapter 1 introduced the goals and specific questions of the New Chance Observational Study. In addition, it presented the context of the New Chance Observational Study by providing an overview of the New Chance Program and by summarizing the results available for the full evaluation of the program. We turn now to focus specifically on the *methodology of the observational study*. We first describe the sample for the observational study, noting how it was selected, presenting sample characteristics, and summarizing impact findings from the 18- and 42-month follow-up surveys specifically for the embedded observational study sample. We then turn to descriptions of the procedures involved in the observational study and of the specific parenting measures available for the observational study sample. We conclude with a summary of the analytic strategy used to address the questions of the New Chance Observational Study.

I. The Sample

A. The Identification of an Initial Pool of Families for the Observational Study

The New Chance Observational Study was "embedded" within the larger New Chance Evaluation in order to provide detailed information about parenting behavior for a subset of New Chance sample members. For the observational study sample, data collected during the observational session was combined with information collected during the 18- and 42-month follow-up surveys of sample members, as well as the participation data about experimental group members from the New Chance Management Information System. Maternal self-reports and interviewer ratings from the follow-up surveys and from a brief interview carried out at the time of the observational session complement data derived from the videotaped interactions of mothers and children during the observational session.

Selection of the observational study sample sought to optimize the analytical value of the information to be collected within practical constraints. Resource constraints precluded including all 16 New Chance sites in such an intensive effort as the observational study (and one requiring costly equipment and props). Therefore, sample selection involved choosing both sites and respondents within sites. This was not entirely a sequential process, since characteristics of sample members in each site played a role in site selection. The potential pool of participants in the observational study was narrowed on the basis of the age of the focal child, ethnicity, and timing of sample enrollment. We examined the potential contribution of each of the 16 New Chance sites to the observational study sample within these constraints. We hoped to limit the number of sites to perhaps five, but found it necessary to include seven sites in order to identify sufficient sample members to yield the target of 300 completed observational sessions.

Although the focal children in the full New Chance Evaluation ranged in age from 18 months to 8 years at the time of the 18-month follow-up, in the observational study we chose to focus on a narrower age range. As described in detail in Chapters 4 and 5 of this monograph, the mother-child interaction tasks used in the observational study had been used in Egeland and Snow's research with preschool-age children. Focusing on children from the New Chance Evaluation who were approximately 30 to 60 months old at the time of the observational session allowed us to use the activities with which the research teams headed by Egeland and Snow had considerable experience and to include many of the focal children from the larger evaluation.

We anticipated wanting to examine parenting behavior in light of ethnic differences within the observational study sample, because of evidence that group differences exist in patterns of mother-child interaction and in the relationship of such patterns to child outcomes (see Garcia Coll, 1990; Bradley et al., 1989). However, it turned out not to be possible to select sites in a way that would result in sufficient representation of the three predominant groups in the full evaluation sample (blacks, Hispanics, and whites). Both practical considerations (such as variation in the sample enrollment schedule across sites) and the intent that this not be a study that examined parenting only among minority families led us to restrict the sample to black and white members of the larger New Chance sample.

Another objective was to conduct each observational session as soon as possible after the 18-month interview, but in no case more than four months later. With data collection for the observational study to begin almost a year after the 18-month anniversary of the first monthly cohort enrolled in the full evaluation, some respondents were already beyond the four-month maximum time period we wished to allow. In addition, sites varied with respect to the number of sample members ultimately enrolled and the speed with which they were enrolled over the two-year sample enrollment period. Thus, another criterion for evaluating sites for inclusion in the observational study was the size of the pool remaining after sample members enrolled during the first nine enrollment months were omitted.

With the potential pool in each site narrowed based on age of the focal child, ethnicity, and timing of sample enrollment, we chose seven sites[1] that would provide sufficient sample members to yield the target of 300 completed sessions. In each site, all of the black and white respondents with an 18-month anniversary date between November 1991 and January 1993 and a focal child who would be 30 to 60 months old at the time of the 18-month interview — or within four months of the interview — were included in a preliminary pool of families.[2] Experimentals constituted 63 percent of this preliminary pool of families and controls 37 percent. Although we had hoped to have a large enough pool to permit oversampling of controls (to provide more cases for analysis), we were not able to modify appreciably the two-to-one experimental-control ratio of the full evaluation sample. In all, 438 New Chance sample members were included in an initial pool of families for inclusion in the observational study, based on the criteria above.

B. Response Rates, Experimental Balance, and Representativeness of the Observational Study Participants

The initial pool of 438 families included 63 sample members who were ultimately considered ineligible for participation in the study because they did not meet all of the study requirements, most notably completion of an 18-month interview with the focal child present.[3] A group of 375 families were then considered eligible for the study. Of this group, 297 actually participated in the observational study, for a response rate of 79 percent, lower than the 91 percent in the 18-month survey in the full evaluation. Twelve percent refused.[4] For the remaining 9 percent of eligible respondents who did not participate, logistical problems in completing the field work were responsible; nearly two-thirds of these cases were simply not completed before the four-month interviewing "window" closed or the child "aged out" of the target age range;[5] others could not be located or had moved from the area.

Of the 297 completed cases, 290 resulted in codable videotapes. Technical difficulties occasionally rendered tapes unusable, as did some of the very early work of a few interviewers. Thus, the analyses described in this monograph are based on 77 percent of the eligible sample. As seen above, some, but not all, of this attrition was based on self-selection.

---

[1] The Bronx, Detroit, Lexington, Harlem, Philadelphia, Pittsburgh, and Portland.

[2] The inclusion of children who had not yet "aged in" to eligibility or were about to "age out" required attention to maintain the integrity of the target age range. Nonetheless, 12 children who participated in the study fell outside the age range, resulting in an actual age range of 27 to 63 months.

[3] In 65 percent of these cases, no 18-month interview was completed at all, typically because the respondent could not be located, while in another 21 percent the randomly selected child was not in the home and there was no other child who could be substituted. In the remaining cases deemed ineligible, either the focal child from the 18-month interview was no longer in the home at the time of the observational session or selection of a substitute child for the 18-month interview or delays in completing that interview resulted in the child being outside the target age range.

[4] Obviously the demands and intrusiveness of the observational session were greater than for a traditional survey interview, no doubt leading some respondents to decline. In addition, concern about the potential intrusiveness of the work led us to counsel interviewers not to press as hard to "convert" those who initially refused as they might in more traditional survey work.

[5] These cases no doubt included an undetermined number of "soft" refusals, that is, those who did not actually refuse but broke appointments, could not find a convenient time for the session, and so on.

The group of 290 cases on which the analyses are based includes 184 experimentals and 106 controls, the same ratio of 63 to 37 percent as in the initial pool of 438 families. Additional analyses were undertaken to assess the comparability of the experimental and control group members who participated in the observational study.

Mothers in the full evaluation sample had been *randomly* assigned to either the experimental or control group; analyses carried out with the full sample confirm that these groups did not differ significantly in terms of baseline characteristics (see Quint, Polit, Bos and Cave, 1994, Appendix A). Parallel analyses were carried out for the observational study sample. Here we focus on the 290 cases actually included in analyses for the New Chance Observational Study, and thus do not include the 7 cases with unusable videotapes.

Regression models were estimated in which the dependent variable was a dummy indicating experimental status (that is, membership in the experimental group versus the control group) and the independent variables were characteristics of the children, mothers, and families at baseline. The resulting F-statistic was not statistically significant, indicating that any difference in baseline characteristics between experimentals and controls in the observational study sample can be attributed to chance (see Appendix Table A.1). Thus, any differences observed between the groups on the measures in this study cannot be attributed to prior group differences.

We noted that self-selection was one factor in the proportion of eligible families who participated in the observational study. In addition, the denominator of cases ultimately determined to be "eligible" may itself have been affected by self-selection (such as failure to participate in the 18-month survey) and characteristics holding potential substantive interest (such as the focal child not living with the mother). Thus, we undertook additional analyses to assess the representativeness of the observational study participants. Here, rather than focusing on the 290 cases included in our data analyses, our sample is the larger group of 297 who actually completed the observational study procedures. We include here the 7 cases that completed the procedures of the observational study but whose videotapes were unusable because the underlying question is whether those families who actually participated in the study differed from those in the initial pool in the seven sites who did not participate.

Accordingly, we compared the baseline characteristics of these 297 families with those of other families in the seven study sites included in the initial pool of families who did not ultimately participate (n = 141). Families who participated in the observational study differed significantly from those in the initial pool of families who did not participate with regard to welfare status at baseline (see Appendix Table B.1). Mothers who completed the observational study procedures were less likely to be receiving welfare at baseline. No other markers of disadvantage at baseline (for example, mothers' education, employment, receipt of welfare as a child) differed across these two groups. These findings point to little difference between those who participated in the New Chance Observational Study and those in the initial pool of families who did not participate beyond what might be expected by chance.[6]

---

[6] These analyses were replicated contrasting the 290 families whose data were actually included in analyses for the New Chance Observational Study with 148 families for whom we do not have data (here including the 7 fami-

(continued)

In sum, these findings indicate that the experimental and control groups included in data analyses for the New Chance Observational Study sample did not differ in terms of measured baseline characteristics in a way other than what might be caused by sampling variation. Further, we see little indication that those families who actually participated in the observational study differed from those included in the initial pool of families at the seven sites who did not participate beyond what might be expected by chance. Therefore, participants in the New Chance Observational Study may be seen as generally representative of the larger pool of families in the seven sites initially considered for inclusion.

It is important to note that while we can view the observational study sample as generally representative of this larger pool of families at the seven sites, *the findings of the observational study should not be seen as generalizing to all of the families in the full New Chance Evaluation.* The larger evaluation sample, for example, included families with focal children in a broader age range, families from additional sites, and families of Hispanic origin.

C. A Profile of the Observational Study Sample

1. Characteristics at Baseline. The observational study subjects were 290 of the young women from the overall New Chance sample and their focal children.[7] The age ranges of the focal children extended slightly beyond the 30-60 month age range originally intended; the actual range was 27-63 months, with 12 children falling outside the planned range. Among the 290 families, 184 were in the experimental group (with 98 male and 86 female focal children), and 106 were in the control group (with 50 male and 56 female focal children). The ages of the focal children averaged 3.6 years at the time of the observational session.

Table 2.1 provides detailed background information about mothers in the observational study sample, drawn from the enrollment information provided just before random assignment. Whereas Chapter 1 presented information to describe the respondents in the full New Chance Evaluation sample, here we focus specifically on the respondents in the embedded observational study.

---

lies with unusable videotapes with the 141 families who were not included from the initial pool for the reasons discussed above). In these analyses two significant differences were found: in addition to being less likely to receive welfare at baseline, families with usable observational data also differed in terms of the highest grade in school the mother had completed. Yet the pattern of findings regarding educational attainment at baseline was complex. While mothers represented in the analysis were somewhat less likely to have completed 11th grade or above (36.2 versus 38.5 percent), and somewhat more likely to have completed 8th grade or below (10.3 versus 8.8 percent), they were also more likely to have completed 10th grade (30.7 versus 25.0 percent) and less likely to have completed 9th grade (22.8 versus 27.7 percent). As for the analysis reported on in Appendix Table B.1, we do not see a clear pattern pointing to greater or lesser disadvantage in the New Chance Observational Study sample included for data analyses.

[7] The "focal" child was the child to whom direct assessments of cognitive development were administered and the child who was the focus of the mother's, interviewer's and teacher's responses to specific survey items concerning parenting and development. When the mother had more than one child at baseline, one was randomly selected to be the focal child. The selection of the focal child occurred after baseline data collection but before the 18-month follow-up.

Table 2.1

Selected Characteristics of the New Chance Observational Study Sample at Baseline

Characteristic	Average or Percent
Demographic Characteristics	
Average age (years)	18.7
Average age at first child's birth (years)	16.6
Race (%)	
Black, non-Hispanic	84.1
White	15.9
Ever married (%)	5.2
Had more than one child (%)	36.9
Average number of pregnancies	1.9
Average age of youngest child (years)	1.4
Education Characteristics	
Received high school diploma or GED (%)	6.6
Average highest grade completed	10.0
Average number of years since last attended school	2.1
Average reading level (grade equivalent)	7.4
Employment and Welfare Receipt	
Had no employment experience (%)	22.8
Did not work in past 12 months (%)	60.7
Average earnings in past 12 months (%)	
$0 - $500	81.7
$501 or more	18.3
Recipient of (%)	
Medicaid	87.8
Food stamps	82.5
Public housing	27.2
Income from a job	4.5
Receiving AFDC on own or other's grant (%)	96.2
Family was ever on AFDC when respondent was growing up (%)	71.0
Living Arrangement	
Living with mother (%)	36.7
Living with father (%)	4.2
Living with husband/boyfriend (%)	10.6
Psychosocial Characteristics	
CES-D (depression) score[a] (%)	
0-15 (not at risk)	48.6
16-23 (at some risk)	26.6
24-60 (at high risk)	24.8
Average Self-Esteem score[b]	39.1

(continued)

Table 2.1 (continued)

	Average or Percent
Average number of sources of emotional support	3.2
Average level of satisfaction with emotional support[c]	4.2
Average Locus of Control score[d]	22.3
Service Receipt	
Services received in the 60 days before random assignment (%)	
Health care for child	87.8
Family planning	24.3
Mental health	1.0
Health care for self	64.6
Parenting instruction	11.1
Life skills training	2.8
Counseling	3.1
Other services	3.1
No services	6.9
Sample size	290

Source: New Chance baseline enrollment data.

Notes: Calculations for this table used data for all 290 respondents for whom there were baseline enrollment data, including those who were randomly assigned to New Chance, but did not participate. The sample size may fall short of the number reported because of missing or unusable items from some respondents' questionnaires.

Distributions may not total to 100.0 percent because of rounding.

[a] The Center for Epidemiological Studies Depression (CES-D) Scale is a widely used measure of depression; scores can range from 0 to 60. Those with scores below 16 on the CES-D are considered to be not at risk of depression; those with scores of 16 and above are considered at risk.

[b] The measure of self-esteem used was the Rosenberg Self-Esteem Scale, a 10-item scale that assesses a person's global sense of self-worth. Scores can range from 10 to 50; 30 is considered the neutral midpoint. A higher number indicates higher self-esteem.

[c] Enrollees were asked about their degree of satisfaction with the emotional support ("people who listen to you, reassure you, and show you they care") they received. Levels range from 1 (very dissatisfied) to 5 (very satisfied).

[d] The Locus of Control Scale is a 6-item adaptation of the longer scale originally developed by Julien Rotter (1966). Scores can range from 6 to 30; 18 is considered the neutral midpoint.

Not unexpectedly, the majority of the New Chance Observational Study sample respondents were characterized by social and economic disadvantage at the time of application to the program. The young women in the sample had given birth on average at 16.6 years, and all by age 19. At baseline, the participants were still relatively young, averaging 18.7 years. The vast majority of New Chance Observational Study respondents, at enrollment, had never married (94.8 percent) and had not earned a GED or high school diploma (93.4 percent), and most had not worked in the prior 12 months (60.7 percent). Nearly three-quarters (71 percent) of sample members' families had received AFDC when the respondent was growing up.

Just over one-third (36.7 percent) of all sample members were living with their mother at the time of random assignment, while 10.6 percent reported living with a husband or boyfriend. In the observational study sample, 51.4 percent of the mothers were at some risk for depression. However, self-esteem and social support scores are not notably low.

At the time of enrollment, applicants reported on services received in the prior 60 days. A nontrivial proportion (11.1 percent) of observational study sample members reported receiving some form of parenting instruction preceding random assignment. This underscores the necessity of considering any program impacts on parenting through the New Chance parenting education classes as occurring over and above the parenting instruction that sample members may have already received (or that control group mothers continued to receive) in the community.

2. 18-Month and 42-Month Follow-Up Findings. Tables 2.2–2.9 present findings from the 18-month and 42-month follow-up surveys for the observational study sample. Although the broad patterning of findings is similar for the full evaluation sample and the observational study sample, there are some differences in the program impact findings. Here we note these similarities and differences. For the earlier follow-up, we include data from all 290 respondents for whom there are 18-month follow-up data as well as data from the observational session. For the 42-month findings, we are limited to the 267 families in the observational study sample for whom there are data from *both* follow-up surveys. Because findings on parenting variables from the 18-month survey are discussed in detail in Chapter 6, they are not summarized here.

18-Month Findings. Tables 2.2–2.9 reveal that program impacts from the 18-month follow-up for the observational study sample were broadly similar to those in the full evaluation sample. Yet a number of specific differences can also be noted. In these instances, usually a difference that was statistically significant in the full evaluation sample did not reach statistical significance in the smaller observational study sample, although the direction of the mean or proportion scores remained the same.

For example, differences in educational attainment for the observational study sample are similar to those for the full evaluation sample: mothers in the experimental group were significantly more likely to have completed the GED, but less likely to have completed high school (Table 2.2). Regarding economic self-sufficiency, in the full evaluation sample, at the 18-month follow-up control group mothers had significantly higher cumulative earnings and there were indications that they had been employed more in the first months after enrollment in the evaluation. In the observational study sample, as can be seen in Table 2.3, means for total earnings and

Table 2.2

Impacts on Educational Attainment, Literacy, and Classes in Education, Training, or Skills at 18-Month and 42-Month Follow-Ups

	18-Month Follow-Up			42-Month Follow-Up		
Outcome	Experimentals	Controls	p[a]	Experimentals	Controls	p[a]
GED or high school diploma (%)	34.1	29.5	.41	42.8	41.9	.89
GED (%)	29.3	17.1	.02 **	37.5	27.3	.09 *
High school diploma (%)	5.4	12.4	.03 **	5.9	14.4	.02 **
Trade certificate or license (%)	11.7	10.1	.83	25.2	27.9	.64
College credit toward degree (%)	7.2	3.6	.22	14.7	10.3	.32
Average TABE reading score at 18 months	745.8	743.8	.58	---	---	---
Reading level at 18 months (%)						
7th grade or below	47.3	49.0	.76	---	---	---
8th or 9th grade	25.0	23.1	.72	---	---	---
10th or 11th grade	13.5	13.6	.97	---	---	---
12th grade or above	14.2	14.3	.97	---	---	---
Ever employed or in education or training or work internship (%)	96.0	79.6	.01 ***	---	---	---
Employed or in education or training or work internship at 18 months (%)	29.7	36.1	.28	---	---	---
Sample size	184	106		170	97	

Sources: New Chance 18-month and 42-month follow-up surveys.

Notes: Calculations for the 18-month follow-up columns used data for all respondents who completed both the observational session and the 18-month follow-up survey. Calculations for the 42-month follow-up columns used data for all respondents who completed the observational session, the 18-month follow-up survey, and the 42-month follow-up survey. Calculations include those with values of zero for outcomes and New Chance enrollees (i.e., experimentals) who did not participate in the program. The sample sizes may fall slightly short of the numbers reported because of missing or unusable items from some respondents' questionnaires.

The averages or percentages are adjusted using linear analysis of covariance procedures controlling for seven kinds of difference in characteristics before random assignment (age of child, maternal literacy, whether mother had more than one child, gender of child, race/ethnicity, Philadelphia site, and Portland site). Distributions may not add to 100.0 percent because of rounding. Dashes indicate unavailable data.

[a] An F-test was applied to each adjusted difference between average experimental and control group outcomes. The column labeled "p" is the statistical significance level of the difference between experimental and control group outcomes. That is, p is the probability that average outcomes are different only because of random error. Statistical significance levels are indicated as ***<= 1 percent, ** <= 5 percent, *<=10 percent.

Table 2.3

Impacts on Time Spent Working and AFDC Receipt at 18-Month and 42-Month Follow-Ups

	18 Month Follow-Up			42-Month Follow-Up		
Outcome	Experimentals	Controls	p[a]	Experimentals	Controls	p[a]
Since random assignment:						
Ever employed (%)	42.0	43.2	.84	69.4	63.0	.29
Total number of hours worked	299.5	309.0	.90	1095.0	1109.0	.45
Total number weeks worked	10.0	10.5	.84	31.5	34.2	.61
Total earnings ($)	1.345	1,522	.64	5,481	6,399	.45
Ever received AFDC (%)	99.5	99.9	.57	99.5	99.9	.61
Ever combined AFDC and work, months 1-42 (%)	--	--	--	65.6	61.4	.50
Average number of months of AFDC receipt	17.0	16.5	.17	38.4	37.3	.30
Receiving AFDC on own grant (%)	86.5	89.5	.45	81.8	76.2	.27
Sample size	184	106		170	97	

Sources: New Chance 18-month and 42-month follow-up surveys.

Notes: Calculations for the 18-month follow-up columns used data for all respondents who completed both the observational session and the 18-month follow-up survey. Calculations for the 42-month follow-up columns used data for all respondents who completed the observational session, the 18-month follow-up survey, and the 42-month follow-up survey. Calculations include those with values of zero for outcomes and New Chance enrollees (i.e., experimentals) who did not participate in the program. The sample sizes may fall slightly short of the numbers reported because of missing or unusable items from some respondents' questionnaires.

The averages or percentages are adjusted using linear analysis of covariance procedures controlling for seven kinds of difference in characteristics before random assignment (age of child, maternal literacy, whether mother had more than one child, gender of child, race/ethnicity, Philadelphia site, and Portland site). Dashes indicate unavailable data.

[a] An F-test was applied to each adjusted difference between average experimental and control group outcomes. The column labeled "p" is the statistical significance level of the difference between experimental and control group outcomes. That is, p is the probability that average outcomes are different only because of random error. Statistical significance levels are indicated as ***<= 1 percent, ** <= 5 percent, *<=10 percent.

Table 2.4

Impacts on Fertility-Related Outcomes at 18-Month and 42-Month Follow-Ups

	18-Month Follow-Up			42-Month Follow-Up		
Outcome	Experimentals	Controls	p[a]	Experimentals	Controls	p[a]
During follow-up, had at least one: (%)						
Pregnancy	59.2	58.0	.31	76.4	77.5	.84
Birth	33.7	32.3	.40	58.0	61.2	.61
Abortion	17.1	9.9	.10 *	22.4	17.4	.33
Number of pregnancies during follow-up (%)						
0	40.1	47.0	.31	55.1	62.3	.26
1	44.6	38.6	.33	37.6	31.0	.28
2 or more	14.6	14.3	.96	7.3	6.8	.86
Birth control status at follow-up (%)						
Sexually abstinent for at least the prior two months, not pregnant	13.8	17.9	.37	16.4	20.6	.41
Sexually active, contracepting regularly	40.0	53.1	.04 **	40.5	43.8	.62
Sexually active, not contracepting regularly	33.1	21.1	.04 **	31.8	23.7	.17
Pregnant at follow-up	12.9	7.9	.20	10.9	12.9	.63
Sample size	184	106		170	97	

Sources: New Chance 18-month and 42-month follow-up surveys.

Notes: Calculations for the 18-month follow-up columns used data for all respondents who completed both the observational session and the 18-month follow-up survey. Calculations for the 42-month follow-up columns used data for all respondents who completed the observational session, the 18-month follow-up survey, and the 42-month follow-up survey. Calculations include those with values of zero for outcomes and New Chance enrollees (i.e., experimentals) who did not participate in the program. The sample sizes may fall slightly short of the numbers reported because of missing or unusable items from some respondents' questionnaires.

The percentages are adjusted using linear analysis of covariance procedures controlling for seven kinds of difference in characteristics before random assignment (age of child, maternal literacy, whether mother had more than one child, gender of child, race/ethnicity, Philadelphia site, and Portland site). Distributions may not add to 100.0 percent because of rounding.

[a] An F-test was applied to each adjusted difference between average experimental and control group outcomes. The column labeled "p" is the statistical significance level of the difference between experimental and control group outcomes. That is, p is the probability that average outcomes are different only because of random error. Statistical significance levels are indicated as ***<= 1 percent, ** <= 5 percent, *<=10 percent.

Table 2.5

Impacts on Marital Status and Residence at 18-Month and 42-Month Follow-Ups

	18-Month Follow-Up			42-Month Follow-Up		
Outcome	Experimentals	Controls	p[a]	Experimentals	Controls	p[a]
Married (%)	9.2	5.8	.27	10.2	5.8	.23
Never married (%)	89.1	92.5	.30	82.2	88.9	.13
Other marital status (%)	1.7	1.7	.97	7.6	5.3	.48
Living with partner or husband (%)	21.3	15.0	.16	24.7	23.8	.36
Living with parent or grandparent (%)	30.8	42.0	.05 **	19.8	24.0	.43
Living with children only (%)	37.3	30.6	.26	41.9	44.1	.72
Living alone or with other (%)	10.8	12.5	.67	13.6	8.1	.19
Number of household members (average)	4.2	4.7	.08 *	4.3	4.3	.97
Sample size	184	106		170	97	

Sources: New Chance 18-month and 42-month follow-up surveys.

Notes: Calculations for the 18-month follow-up columns used data for all respondents who completed both the observational session and the 18-month follow-up survey. Calculations for the 42-month follow-up columns used data for all respondents who completed the observational session, the 18-month follow-up survey, and the 42-month follow-up survey. Calculations include those with values of zero for outcomes and New Chance enrollees (i.e., experimentals) who did not participate in the program. The sample sizes may fall slightly short of the numbers reported because of missing or unusable items from some respondents' questionnaires.

The averages or percentages are adjusted using linear analysis of covariance procedures controlling for seven kinds of difference in characteristics before random assignment (age of child, maternal literacy, whether mother had more than one child, gender of child, race/ethnicity, Philadelphia site, and Portland site).

[a] An F-test was applied to each adjusted difference between average experimental and control group outcomes. The column labeled "p" is the statistical significance level of the difference between experimental and control group outcomes. That is, p is the probability that average outcomes are different only because of random error. Statistical significance levels are indicated as ***<= 1 percent, ** <= 5 percent, *<= 10 percent.

Table 2.6

Impacts on Child Care Use for the Focal Child at or Within 18-Month and 42- Month Follow-Ups

	18-Month Follow-Up			42-Month Follow-Up		
Outcome	Experimentals	Controls	p[a]	Experimentals	Controls	p[a]
Ever in regular child care arrangement during months 1-18 (%)	94.6	83.0	.00 ***	---	---	---
Ever in child care arrangement before age 1 (%)	31.0	30.1	.91	---	---	---
Between baseline and 18-month follow-up, use of/attendance in: (%)						
School	1.2	.70	.67	---	---	---
Head Start	15.0	12.7	.57	---	---	---
Day care center/preschool	64.1	30.3	.00 ***	---	---	---
Family day care/unrelated sitter	24.2	22.2	.71	---	---	---
Care by grandparent	43.3	42.7	.92	---	---	---
Care by another relative	21.4	24.1	.61	---	---	---
Care by husband/partner	21.2	18.8	.63	---	---	---
Average number of months from baseline to 18-month follow-up that child was in/used:						
Day care center/preschool	4.5	2.2	.00 ***	---	---	---
Family day care/unrelated sitter	1.4	1.4	.88	---	---	---
Care by grandparent	2.7	3.3	.42	---	---	---
At time of interviews, use of/attendance in: (%)						
Child care/school	48.6	52.5	.52	91.3	91.3	.95
Day care center or preschool	18.8	18.2	.90	13.7	13.0	.87
Regular care by grandparent	15.3	24.2	.07 *	30.4	30.6	.97
By 42-month interview, use of: (%)						
Day care center/preschool	---	---	---	58.9	52.4	.31
Head Start program	---	---	---	49.8	58.3	.20
Sample size	184	106		170	97	

Sources: New Chance 18-month and 42-month follow-up surveys.

Notes: Calculations for the 18-month follow-up columns used data for all respondents who completed both the observational session and the 18-month follow-up survey. Calculations for the 42-month follow-up columns used data for all respondents who completed the observational session, the 18-month follow-up survey, and the 42-month follow-up survey. Calculations include those with values of zero for outcomes and New Chance enrollees (i.e., experimentals) who did not participate in the program. The sample sizes may fall slightly short of the numbers reported because of missing or unusable items from some respondents' questionnaires.

The averages or percentages are adjusted using linear analysis of covariance procedures controlling for seven kinds of difference in characteristics before random assignment (age of child, maternal literacy, whether mother had more than one child, gender of child, race/ethnicity, Philadelphia site, and Portland site). Dashes indicate unavailable data.

[a] An F-test was applied to each adjusted difference between average experimental and control group outcomes. The column labeled "p" is the statistical significance level of the difference between experimental and control group outcomes. That is, p is the probability that average outcomes are different only because of random error. Statistical significance levels are indicated as ***<= 1 percent, ** <= 5 percent, *<=10 percent.

Table 2.7

Impacts on Psychological Characteristics of Respondent at 18-Month and 42-Month Follow-Ups

	18-Month Follow-Up			42-Month Follow-Up		
Outcome	Experimentals	Controls	p[a]	Experimentals	Controls	p[a]
Average CES-D (depression) score[b]	15.3	13.4	.09 *	15.3	15.5	.85
Depressed at follow-up (%)	40.7	36.0	.44	41.1	44.6	.61
Felt stressed all or much of the time in past month (%)	---	---	---	38.7	28.1	.09 *
Average Mastery score[c]	22.2	22.4	.70	22.4	22.6	.62
Difficult Life Circumstances score[d]	2.8	2.7	.50	2.8	2.6	.31
Number of social supports	2.0	2.0	.90	1.7	1.7	.61
No social support at follow-up (%)	7.0	4.9	.47	6.0	6.0	.85
Satisfaction with emotional support[e]	8.1	8.0	.81	8.3	8.6	.26
Sample size	184	106		170	97	

Sources: New Chance 18-month and 42-month follow-up surveys.

Notes: Calculations for the 18-month follow-up columns used data for all respondents who completed both the observational session and the 18-month follow-up survey. Calculations for the 42-month follow-up columns used data for all respondents who completed the observational session, the 18-month follow-up survey, and the 42-month follow-up survey. Calculations include those with values of zero for outcomes and New Chance enrollees (i.e., experimentals) who did not participate in the program. The sample sizes may fall slightly short of the numbers reported because of missing or unusable items from some respondents' questionnaires.

The averages or percentages are adjusted using linear analysis of covariance procedures controlling for seven kinds of difference in characteristics before random assignment (age of child, maternal literacy, whether mother had more than one child, gender of child, race/ethnicity, Philadelphia site, and Portland site). Dashes indicate unavailable data.

[a] An F-test was applied to each adjusted difference between average experimental and control group outcomes. The column labeled "p" is the statistical significance level of the difference between experimental and control group outcomes. That is, p is the probability that average outcomes are different only because of random error. Statistical significance levels are indicated as ***<= 1 percent, ** <= 5 percent, *<=10 percent.

[b] The Center for Epidemiological Studies Depression (CES-D) Scale is a widely used measure of depression; scores can range from 0 to 60.

[c] The Mastery Scale measures a sense of mastery over personal events. Scores can range from 7 to 28.

[d] The Difficult Life Circumstances scores represent the total number of ongoing problems or stressors the respondent faces, out of a list of 10 problems.

[e] Satisfaction with emotional support was rated on a scale from 0 ("extremely dissatisfied") to 10 ("extremely satisfied").

Table 2.8

Impacts on Child Development Outcomes at 42-Month Follow-Up

Outcome	Experimentals	Controls	Difference	p[a]
Cognitive				
Standard score on School Readiness Component of Bracken Basic Concept Scale	7.0	7.4	-0.4	.34
Socioemotional				
Behavior Problems Index (BPI)				
Total BPI scale, standard score	108.7	106.9	1.8	.32
BPI subscale scores				
Antisocial	110.2	108.9	1.2	.52
Anxious/depressed	106.0	104.5	1.6	.28
Headstrong	100.7	99.1	1.6	.30
Hyperactive	107.8	107.3	0.5	.79
Dependent	107.5	107.6	-0.1	.96
Peer conflict/withdrawn	107.6	105.0	2.6	.15
Positive Behavior Index (PBI)				
Total PBI score	193.8	202.8	-9.0	.04 **
PBI subscale scores				
Compliance	62.5	65.8	-3.3	.09 *
Social competence	88.3	92.4	-4.2	.04 **
Autonomy	43.1	44.3	-1.2	.15
Sample size	170	97		

Source: New Chance 42-month follow-up survey.

Notes: Calculations for this table used data for all respondents who completed the observational session, the 18-month follow-up survey, and the 42-month follow-up survey. Calculations include those with values of zero for outcomes and New Chance enrollees (i.e., experimentals) who did not participate in the program. The sample sizes may fall slightly short of the numbers reported because of missing or unusable items from some respondents' questionnaires.

The averages are adjusted using linear analysis of covariance procedures controlling for seven kinds of difference in characteristics before random assignment (age of child, maternal literacy, whether mother had more than one child, gender of child, race/ethnicity, Philadelphia site, and Portland site).

[a] An F-test was applied to each adjusted difference between average experimental and control group outcomes. The column labeled "p" is the statistical significance level of the difference between experimental and control group outcomes. That is, p is the probability that average outcomes are different only because of random error. Statistical significance levels are indicated as ***<= 1 percent, ** <= 5 percent, *<=10 percent.

Table 2.9

Impacts on Child Development Outcomes at 42-Month Follow-Up, Teacher Survey Subsample

Outcome	Experimentals	Controls	Difference	p[a]
Behavior Problems Index				
Raw total score, teacher survey	7.8	7.7	0.1	.93
Raw total score, maternal report	9.6	9.0	0.6	.53
Standardized total score, maternal report	108.5	107.2	1.3	.58
Positive Behavior Index				
Raw total score, teacher survey	171.3	173.6	-2.3	.80
Raw total score, maternal report	196.8	204.4	-7.6	.18
Average teacher rating (1-5) of child's overall academic performance	3.1	3.0	0.1	.64
Average maternal rating (0-10) of child's academic performance	8.4	8.8	-0.4	.24
Teacher's report of disciplinary problems (%)				
Has taken disciplinary action more than once a week	29.2	24.2	5.0	.49
Has had behavior problem requiring parental notification	26.9	22.5	4.4	.53
Has had multiple behavior problems requiring parental notification	15.2	9.4	5.8	.29
Mother's report of school notification of behavior problem (%)	19.5	10.8	8.8	.21
Teacher rated child as worse than other students in class in terms of (%)				
Getting along with other students	18.7	23.7	-5.0	.46
Getting along with teachers	13.8	11.9	1.9	.74
His/her self-esteem	24.0	23.3	0.7	.92
His/her motivation to do well	27.7	24.3	3.4	.64
His/her overall school adjustment	21.6	19.1	2.5	.71
Teacher reported meeting with mother for a general discussion of child (%)	84.6	78.2	6.3	.34
Sample size	94	62		

Sources: New Chance 42-month teacher and maternal report surveys.

Notes: Calculations for this table used data for all respondents who completed the observational session, the 18-month follow-up survey, the 42-month follow-up survey, and for whom there was a completed teacher survey. Calculations include those with values of zero for outcomes and New Chance enrollees (i.e., experimentals) who did not participate in the program. The sample sizes may fall slightly short of the numbers reported because of missing or unusable items from some respondents' questionnaires.

The averages or percentages are adjusted using linear analysis of covariance procedures controlling for seven kinds of difference in characteristics before random assignment (age of child, maternal literacy, whether mother had more than one child, gender of child, race/ethnicity, Philadelphia site, and Portland site).

[a]An F-test was applied to each adjusted difference between average experimental and control group outcomes. The column labeled "p" is the statistical significance level of the difference between experimental and control group outcomes. That is, p is the probability that average outcomes are different only because of random error. Statistical significance levels are indicated as ***<= 1 percent, ** <= 5 percent, *<=10 percent.

for the duration of employment variables are in the same direction, but differences do not reach statistical significance.

On fertility measures, within the observational study sample, as in the full evaluation sample, mothers in the experimental group were significantly more likely to report being sexually active but not contracepting regularly, and also significantly more likely to report having had an abortion. While in the full evaluation sample as reported in the 18-month report, experimental group mothers reported more pregnancies than controls during the 18-month follow-up period, this difference did not attain statistical significance in the observational study sample. Table 2.4, however, shows that the proportion of experimental and control group mothers reporting a pregnancy in the observational study sample did follow the same pattern as in the full evaluation sample: 59 percent of experimental group mothers and 58 percent of control group mothers reported a pregnancy.

Regarding residence, Table 2.5 shows a pattern for the observational study sample that parallels that found in the full evaluation sample: a numerically larger proportion of experimental group mothers were living with a partner or husband at the time of the 18-month follow-up, and a numerically smaller proportion were living with a parent or grandparent. While both of these differences were significant in the full evaluation sample, only the difference in residence with a parent or grandparent differed significantly for the observational study sample.

As in the full evaluation sample, children in the observational study sample were significantly more likely to have experienced nonmaternal care, specifically day care or preschool, during the 18-month follow-up period (Table 2.6). Although children in the full evaluation sample were significantly more likely to have experienced child care before their first birthday, there is no parallel significant group difference for the observational study sample.

When we turn to variables reflecting maternal psychological well-being (Table 2.7), however, we see some differences in the pattern of findings for the full evaluation and observational study samples. In the full evaluation sample, experimental group mothers were significantly less likely than control group mothers to report that they had no social support, and reported significantly greater satisfaction with social support. In the observational study sample we see no indication of such a pattern. Mean scores on the variable reflecting satisfaction with social support are nearly identical for the experimental and control in the observational study sample (8.1 and 8.0 in the experimental and control groups, respectively), and the proportion of mothers reporting that they have no social support is numerically (though not significantly) smaller in the control group than in the experimental group (7.0 percent in the experimental group and 4.9 percent in the control group). Further, in the observational study sample, though not in the full evaluation sample, mothers in the experimental group showed significantly more depressive symptomatology than mothers in the control group at the time of the 18-month follow-up. Within the full evaluation sample, mean scores for depression were higher in the experimental group, but only slightly so (mean scores were 16.3 and 15.7 in the experimental and control groups, respectively). In this area, then, at the 18-month follow-up findings diverge somewhat for the observational study sample.

42-Month Findings. We turn now to the question of whether the pattern of program impacts found for the full evaluation sample at 42 months is sustained in the observational study sample. As can be seen in Table 2.2, findings for educational attainment at 42 months in the observational study sample are similar to those for the full evaluation sample: a significantly higher proportion of experimental group mothers had completed the GED, but a lower proportion had completed high school. In the full evaluation sample, a further significant impact reflected the fact that more experimental group mothers had completed some college credits. While the scores for the experimental and control groups in the observational study sample fall in this direction (with 14.7 percent of experimental group mothers and 10.3 percent of control group mothers having completed some college credits), the difference is not statistically significant.

Table 2.3 shows that on several key markers of economic self-sufficiency, findings for the observational study sample parallel those in the full evaluation sample. Specifically, we see no program impacts suggesting that the intervention increased earnings or employment or reduced the proportion of mothers currently receiving AFDC. While in the full evaluation sample a significantly higher proportion of experimental group mothers had combined employment and AFDC receipt at some point, no group difference was found on this variable for the observational study sample (though again, scores fall in the same direction as for the full evaluation sample, with 65.6 percent of experimental group mothers and 61.4 percent of control group mothers having ever combined employment and AFDC receipt during the 42-month follow-up period).

Table 2.4 shows that in the observational study sample, as in the full evaluation sample, no group differences were found at the 42-month follow-up on measures of mothers' fertility or fertility behavior. Regarding residence patterns, however, two differences were noted in the full evaluation sample that are not replicated in the observational study sample (see Table 2.5). In the observational study sample we see no indication of a higher average number of household members in the experimental group. Further, we do not see a significant difference in the observational sample in the proportion of mothers living alone, with friends, or in an institution (though scores for the observational study sample are in the same direction as in the full evaluation sample, with 13.6 percent of experimental group mothers, and 8.1 percent of control group mothers living in this arrangement).

Regarding child care, there was no indication in either the full evaluation sample or the observational study sample (Table 2.6) of a significant group difference in the proportion of families using a child care arrangement *at the time of the 42-month interview*. However, looking *across the entire 42-month follow-up period*, in the full evaluation sample a significantly higher proportion of experimental group mothers reported the use of day care or preschool. In the observational study sample the scores follow this pattern (with 58.9 percent of experimental group mothers and 52.4 percent of control group mothers reporting the use of child care or preschool at some point during the 42-month period), but the difference is not statistically significant.

Regarding mothers' psychological well-being, in the observational study sample (Table 2.7), as in the full evaluation sample, significantly more experimental group mothers reported feeling stressed all or much of the time in the month prior to the interview. However, findings for maternal depression for the observational study sample do not parallel those found for the full evaluation sample. Whereas in the full evaluation sample experimental group mothers had higher mean depression scores, no such difference is detected in the observational study sample, and indeed a slightly (though not significantly) higher proportion of control group mothers than experimental group mothers showed signs of depression at 42 months. We noted above in our summary of 18-month follow-up findings that within the observational study sample (though not in the full evaluation sample) there was a group difference in maternal depression at this earlier time, with experimental group mothers showing more depression. It appears that greater depression among experimental group mothers surfaces in the observational study sample as in the full evaluation sample, only at a somewhat earlier time.

Table 2.8 presents 42-month child outcome findings for the observational study sample. As can be seen, there are two scales of children's social behavior: the Behavior Problems Index (with a total score and six subscale scores) and the Positive Behavior Index (with a total score and three subscale scores). Each of these scales reflects the mother's perceptions of her child's social behavior. In Chapter 1 we noted that in the full evaluation sample, mothers in the experimental group perceived their children's social behavior more negatively in terms of total scores and specific subscale scores on *both* the Behavior Problems Index and the Positive Behavior Index. However, Table 2.8 shows that in the observational study sample, while we still see indications that mothers in the experimental group perceived their children's social behavior more negatively, the negative findings are limited to the Positive Behavior Index only. Mothers in the experimental group described their children as showing less positive social behavior overall, and they perceived differences specifically in the areas of compliance behavior and social competence.

Teachers provided information about 156 of the 290 focal children in the observational study (Table 2.9). For those children covered by the teacher survey, teachers completed the same two scales of children's social behavior as had the mothers: the Behavior Problems Index and the Positive Behavior Index. In addition, teachers completed a series of ratings concerning the children's academic and behavioral adjustment to school. We see no indication within the observational study sample that teachers perceived differences according to group (experimental or control) in the children's social behavior (as measured by the Behavior Problems Index and the Positive Behavior Index) or on ratings of the children's academic or social behavior in school. This parallels the pattern for the full evaluation sample, where teachers reported no overall group difference on the social behavior scales or the ratings of academic and behavioral adjustment to school. Mothers were also asked to rate their children's academic performance and to indicate whether the school had notified them of any behavior problems in school. In the full evaluation sample, as noted in Chapter 1, mothers in the experimental group rated their children's adjustment to school less favorably than mothers in the control group. Within the observational sample, differences on these ratings fell in the same direction, but the differences did not reach statistical significance. For example, as shown in Table 2.9, for the observational study sample, 19.5 percent of mothers in the experimental group and 10.8 percent of those in the control group reported that they had been notified by the school of a behavior problem, a difference that was not statistically significant.

Summary of Findings. In broad terms, we find that the results for the observational study sample largely parallel those in the full evaluation sample. However, in a few noteworthy instances the findings diverge. Of particular importance from the point of view of interpretation of the observational study findings, we see some indications that greater psychological distress within the experimental group surfaced earlier in the observational study sample than in the full evaluation sample. At 18 months, a significant group difference in depression was already apparent in the observational study sample (though not in the full evaluation sample). At 42 months, only one marker of psychological distress (the proportion of mothers reporting feeling stressed all or much of the time in the previous month) was elevated in the observational study sample experimental group while there was a group difference in depression in the full evaluation sample. The 42-month child outcome findings for the observational study sample should be viewed in light of this difference in timing particularly regarding maternal depression. Maternal psychological well-being may play a role in how mothers perceive their children's behavior. For example, greater depression could alter patience and irritability with child behavior, and thus the same child behavior could be perceived differently by mothers who are or are not depressed.

A second difference in results for the full evaluation sample and embedded study sample that will be important to bear in mind concerns child outcomes. While experimental group mothers in the observational study sample, as in the full evaluation sample, perceived their children's social behavior less favorably, this difference did not extend to as many measures within the observational study sample. Experimental group mothers in the observational study sample perceived fewer positive social behaviors in their children, but did not perceive a greater incidence of problem behaviors. Further, among those participating in the teacher survey, there were no group differences in mothers' perception of academic or behavioral adjustment to school within the observational study sample, although in the full sample such perceptions were significantly less favorable among experimental group mothers. The possibility exists that these differences in the findings for the observational study sample in child outcome and maternal psychological well-being variables at 42 months are linked.

3. Findings from the "Brief Interview" Completed at the Time of the Observational Session. A brief interview was completed at the time of the visit to families' homes to videotape mother-child interaction for the observational study. Because the observational session followed the 18-month follow-up (occurring on average 21 months after families' enrollment in the New Chance Evaluation)[8] findings from this brief interview somewhat extend the picture of the families' experiences. Again, because findings for measures of *parenting* obtained in the context of interviews are presented in detail in Chapter 6, our presentation here will not include findings for the parenting Time Use variables from this brief interview. Table 2.10 provides a summary of impacts of the New Chance Program for other variables from the interview carried out as part of the observational study data collection effort (which we will refer to as the "brief interview").

A greater percentage of controls than experimentals were enrolled in GED or basic skills training at the time of the brief interview. Beyond this, there were no further differences across the groups in terms of education or training, employment, or welfare receipt. In terms of residence, a notably greater proportion of experimentals (26 percent) than controls (18 percent) were

---

[8] The extra visit to families' homes carried out to collect data for the New Chance Observational Study occurred on average 21 months after mothers enrollment in the evaluation, but ranged from 16 to 36 months. These visits were scheduled to occur as soon as possible after the 18-month follow-up, and within four months thereafter.

Table 2.10

Impacts on Observational Interview Variables at 21-Month Follow-Up

Variable	Experimentals	Controls	p[a]
School/Day Care Attendance for Focal Child[b]			
Kindergarten/elementary school (%)	2.3	0.8	.34
Head Start (%)	10.1	10.2	.97
Day care or preschool (%)	21.4	21.8	.93
Any school or day care (%)	31.6	29.6	.72
Informal Child Care Use[c]			
Family day care or unrelated sitter (%)	20.2	21.1	.86
Care by respondent's mother (%)	36.7	38.2	.80
Care by another relative (e.g., respondent's sister or aunt) (%)	34.9	40.0	.40
Care by current or former boyfriend/husband (%)	20.5	25.5	.34
Care by respondent's mother or other relative (e.g., sister, aunt)	53.5	62.0	.15
Any informal child care (%)	68.6	77.5	.09 *
Hours per week in all child care arrangements, including school (%)	18.8	19.5	.78
Residence			
Living with husband or boyfriend (%)	26.0	18.2	.10 *
Living with mother or father (%)	23.6	35.4	.03 **
Living with other relatives (%)	32.1	39.5	.18
Average number of people in household (besides respondent and her children)	3.1	3.4	.39
Average number of children respondent gave birth to who live with her now	1.8	1.8	.96
Enrollment/Hours in Education or Training Activity			
High school or college (%)	7.9	11.8	.27
GED or basic skills (%)	7.3	13.7	.08 *
Job training (%)	13.1	13.2	.98
Job search classes (%)	3.1	4.1	.67
Enrolled in any education or training (%)	27.7	33.1	.35
Average hours per week in education/training activity	6.3	7.0	.64
Employment and Welfare Receipt			
Currently working at a paid job (%)	12.5	16.9	.32
Hours per week usually working	3.3	5.1	.18
Currently looking for a job (%)	44.2	43.0	.84
Receiving welfare on own or other's grant (%)	90.0	90.0	.99
Hours per week in education/training/working	9.5	12.0	.19
Other			
Respondent reports that a woman other than herself is like a mother to focal child	64.3	70.8	.27
Life Satisfaction[d]	56.9	60.7	.02 **
Sample size	184	106	

(continued)

BEST COPY AVAILABLE

Table 2.10 (continued)

Source: New Chance brief interview accompanying the observational session.

Notes: Calculations for this table used data for all 290 respondents for whom there were 18-month and 21-month follow-up survey data, including those with values of zero for outcomes and New Chance enrollees (i.e., experimentals) who did not participate in the program. The sample sizes may fall slightly short of the numbers reported because of missing or unusable items from some respondents' questionnaires.

The averages or percentages are adjusted using linear analysis of covariance procedures controlling for seven kinds of difference in characteristics before random assignment (age of child, maternal literacy, whether mother had more than one child, gender of child, race/ethnicity, Philadelphia site, and Portland site).

[a] An F-test was applied to each adjusted difference between average experimental and control group outcomes. The column labeled "p" is the statistical significance level of the difference between experimental and control group outcomes. That is, p is the probability that average outcomes are different only because of random error. Statistical significance levels are indicated as ***<= 1 percent, ** <= 5 percent, *<=10 percent.

[b] When a sample member had more than one child, her response refers to the randomly selected focal child.

[c] Respondents were permitted to report more than one type of child care and schooling; thus, percentages do not sum to 100.

[d] Life Satisfaction was measured using 9 items that asked about both global satisfaction and satisfaction with specific areas of one's life (e.g., "the amount of money you have to live on"). The measure had a coefficient alpha of .76.

living with a husband or boyfriend at the time of the brief interview, while a markedly larger share of controls (35 percent) than experimentals (24 percent) were living with their mother or father only. There were no group differences according to the average number of people in the household.

It is important to note that in the brief interview experimental group mothers in the observational study sample reported themselves to be significantly less satisfied with their lives than controls, according to a 9-item measure of overall Life Satisfaction. There were no other aspects of maternal psychological well-being in the brief interview that could be analyzed.[9] However, as we noted, at 18 months mothers in the observational study sample also showed more depressive symptomatology than mothers in the control group. The finding for the Life Satisfaction measure is in keeping with our portrayal of an earlier surfacing of psychological distress among experimental group mothers in the observational study sample. It will be important to see if impacts on parenting behavior within the observational study sample co-occur with these indications of greater psychological distress among the experimental group mothers.

Although there were no statistically significant differences in attendance at child care centers across the two groups at the time of the brief interview, a greater proportion of children whose mothers were in the control group attended informal care than their counterparts whose mothers were in the experimental group — 78 versus 69 percent.

Having provided an overview of program impacts for the observational study sample (on variables other than parenting measures) from the 18-month and 42-month follow-up surveys and the brief interview that was carried out as part of the observational study, we now turn to a discussion of how the observational study was conducted.

II. Study Procedures

To examine the effects of the New Chance Intervention on parenting behavior within the observational study sample, this report draws on several kinds of information. A primary source, of course, is the videotaped session in which mother-child interaction was observed. In addition, mothers provided self-report data and interviewers recorded ratings from which measures of parenting practices and the home environment were derived. Procedures for coding the individual variables, discussions of their use in previous literature, and reliability statistics for each of the parenting measures are described in detail in Chapters 4–6.

A. The Observational Session

The observational sessions were conducted by teams consisting of a survey interviewer and a videographer. Once the observational study was launched, the mothers were generally told about the study at the time of their 18-month follow-up interview, and the extra visit was scheduled as soon after the interview as possible. If the 18-month interview had already been completed and/or had been administered by a different interviewer when the observational study was launched, a special contact by the observational study interviewer was necessary to arrange for the session. In either case, the interviewer explained about a "special study of mothers and children" that would involve the mother showing her child a book and some toys, brought by the interviewer, while being videotaped.[10]

---

[9] A 5-item measure of depression included in the brief interview at 21 months showed poor reliability and thus was not used in analyses.

Upon arrival at the family's home, the interviewer again explained what the study would entail and obtained written informed consent. The interviewer and videographer then identified with the respondent a suitable place in the home to conduct the session and, together, made any necessary modifications (such as moving furniture). While the videographer set up the taping equipment, the interviewer established rapport with the respondent and conducted the first segment of the mother's interview, focusing on how the mother and child had spent the previous weekday. Respondents used an open-ended format to describe their days, with interviewers probing to understand the nature and duration of each activity and recording activities on a grid marked off in half-hour intervals. When this phase of the session had been completed, the group moved into the observational part of the session itself.

At this time the videographer began the taping, while the interviewer introduced the activities to the mother, following a written "script." (The process of adapting the observational protocol for administration by survey interviewers, including the development of the script, are described in detail in Chapter 11.) Ideally, the interviewer was to do this without the focal child immediately present. If instead it was necessary to distract the child while the interviewer instructed the mother, the videographer entertained the child (and sometimes other children) with a hand puppet. The interviewer was instructed to do everything possible to make the mother feel comfortable and to make sure that she understood the objectives of each activity, while strictly adhering to the protocol. (Data on mothers' subjective reactions to the observational study procedures, and interviewer adherence to the protocol, are presented in Chapter 12.) After completing the explanation of the tasks, the interviewer asked the mother to have the child join them.

During the mother-child interaction phase of the session, the interviewer continued to follow the script, presenting (and removing) the various study materials, generally with the assistance of the videographer,[11] and briefly reminding the mother of the objectives of each new task as she introduced it. Otherwise, she was to observe politely, letting the mother-child interaction unfold on its own without directing it in any way. As outlined below, this part of the session included a book reading activity, lasting 5 to 10 minutes; a series of teaching tasks, lasting 15 to 20 minutes; and a segment in which the interviewer gave the mother a gift to present to her child, lasting about 2 minutes.

- Book reading activity (adapted from previous research by Snow and colleagues; see, for example, Snow et al., 1991). The mother was to read and discuss a children's book, *The Very Hungry Caterpillar* (Carle, 1969), with her child "the way you would usually read or look at a book together."

---

[10] Respondents were advised that they would not need to "do anything special to prepare for the videotaping."

[11] However, videographers were sometimes needed to entertain other children in the household while the mother

(continued)

- Teaching tasks (adapted from previous research by Egeland and colleagues; see, for example, Sroufe, Egeland, and Kreutzer, 1990).[12]
    - Blocks. The mother was to try to get her child to replicate a model by putting together smaller blocks.
    - Wheels. The mother was to try to get her child to name as many things that have wheels as he or she could.
    - Sorting. Depending on the age of the child, and in some cases the child's performance, the mother presented one or two sorting tasks; an easier task that required the child to sort chips of different colors on a board and a more complex task that required the child sort the chips by color, shape, and size.
    - Etch-a-Sketch. On this task, the mother was to try to get her child to draw a line on an Etch-a-Sketch screen that moved from point A to point B on a maze drawn on the screen, but without crossing the lines on the maze.
- Gift interaction (developed by the observational studies team). The mother was presented with a wrapped gift — a kaleidoscope — to give to the child, and the dyad then spent a few minutes interacting around the gift.

The videotaping equipment was turned off at the end of the gift segment. Following the observational session, the second section of the brief interview with the mother was completed, with some of the items completed by the mother in a self-administered questionnaire. This part of the interview included questions concerning mother's current residence, markers of the mother's subjective well-being, information regarding participation in educational or employment activities, and the child's current child care arrangements.

B. The 18-Month Follow-Up

Eighteen months after random assignment a survey was conducted for the full evaluation sample (including the observational study sample) to assess the short-term impacts of the New Chance Intervention. Interviews focused primarily on mothers' education, employment, receipt of benefits, changes in household composition, fertility, parenting, use of child care, and maternal subjective well-being. A literacy test was also administered to the mothers. Interviews typically took place in the respondent's home and lasted about 90 minutes. Maternal reports of parenting behavior were supplemented by interviewers' ratings of the home environment and of the spontaneous interactions of the mother and focal child, who was expected to be present during the interview.

---

and child were working together and so were not always able to assist the interviewer.

[12] The teaching tasks were originally designed by Block and Block for use in their longitudinal study of personality development (Harrington, Block, and Block, 1978). The tasks were modified by Egeland for use in the Mother-Child Project. Props used in the study were designed by Egeland, adapted with input from the observational study team, and constructed at the University of Minnesota.

C. Specific Parenting Measures in the New Chance Observational Study

Table 2.11 lists the specific parenting measures available for the New Chance Observational Study sample, the procedure from which each measure was derived, and the location in this monograph where a detailed discussion of each measure's source and derivation, history, and psychometric reliability can be found.

III. Strategies for Analyses in This Monograph

The results portions of Chapters 4–6 (which focus on the observational measures of affective quality, the literacy-related observational measures, and the interview-based measures of parenting, respectively) begin with a descriptive breakdown of the specific parenting behaviors of interest within the New Chance Observational Study sample — that is, an examination of variation in the parenting behaviors of interest according to mothers' characteristics at baseline and at the 18-month follow-up interview. These descriptive analyses allow us to examine the ways in which parenting behaviors in this sample vary according to the characteristics of the mothers, children, and families at baseline and at the first follow-up. In general, chapters discuss significant subgroup differences only when they occur across two or more of the parenting variables being examined in the chapter. For example, if parenting behavior is found to differ significantly according to a baseline variable such as mother's age at the birth of her first child (ages 13–16 versus 17–19), such differences, while always presented in the relevant table, would be summarized in the text only if they occurred for two or more of the parenting variables considered in that chapter. An exception is made for variables of key substantive or policy interest, such as mother's baseline literacy score in Chapter 5 (which considers maternal behavior relevant to the development of literacy in the child). While the specific placement and numbering of descriptive tables varies throughout Chapters 4–6, each of the results chapters examines parenting behavior in light of baseline characteristics, and analyzes parenting behavior in relation to variables drawn from the 18-month survey. These descriptive analyses address the first set of questions that this monograph seeks to answer:

- **How can we describe the parenting behavior of the mothers in this sample? Is parenting behavior systematically related to family background characteristics? Are there identifiable subgroups of mothers who appear to be at greater risk or, alternately, to show more positive parenting behavior?**

To address this first question, the aim is to describe characteristics of the overall sample of mothers, regardless of their research group status. For this reason, respondents' experimental/control status has been statistically controlled in the initial descriptive analyses presented in Chapters 4–6.

Next, we examine whether participation in the New Chance Program significantly affects key maternal parenting behaviors. In assessing this possibility, we separate the sample into experimentals and controls and determine whether mean scores on parenting variables are statisti-

Table 2.11

Sources of Parenting Measures Included in the New Chance Observational Study

		Time of Collection in New Chance Study			
Type of Parenting Measure	Variables	Observational Session	18-Month Interview	21-Month Interview	Chapter
Observational Measures of Mother-Child Interaction Related to Literacy	*Book Reading Task:*	Derived from two tasks only: book reading task and wheels task			5
	Total Number of Utterances				
	Number of Nonimmediate Utterances				
	Percentage of Immediate Utterances				
	Number of Discussion Topics				
	Book Reading Quality				
	Wheels Task:				
	Objects Named				
	Objects/Elicitations				
	Mother's Ease of Ideas				
Observational Measures of Affective Quality of Mother-Child Interaction	*Ratings of Mother:*	Derived from full videotape			4
	Supportive Presence				
	Intrusiveness				
	Hostility				
	Quality of Instruction				
	Confidence				
	Harsh Treatment				
	Ratings of Child:				
	Persistence				
	Enthusiasm				
	Negativity				
	Compliance				
	Experience of Session				
	Affection to Mother				
	Avoidance of Mother				
	Ratings of Dyadic Behavior:				
	Quality of Relationship				
	Boundary Dissolution				
Combination of Interviewer Ratings and Maternal Report: Home Observation for Measurement of the Environment — Short Form (HOME-SF)	Emotional Support		X		6
	Cognitive Stimulation				
	Physical Environment				
	Harsh Discipline				
	HOME Total				
Maternal Report Scales	Warmth		X		6
	Control		X		
	Stress		X		
	Overall Parenting Time			X	
	Parenting Chore Time			X	

cally different across the two groups. This examination, also presented in Chapters 4–6, is designed to address the second question of the monograph:

- **Did New Chance have positive impacts on mother-child interactions and parenting?**

A statistically significant difference between experimentals and controls indicates that the New Chance Program had a detectable *impact* on the particular behavior of interest.

Once we ascertain the aspects of maternal behavior affected by the New Chance Program, we examine impacts for key subgroups of mothers, defined by their baseline characteristics (and thus unaffected by the program at that point in time). For example, we ask whether a particular parenting impact occurred for subgroups in which the mother had and had not received welfare as a child; in which the mother was or was not living with her parents prior to enrollment in the program. This examination is designed to address the question:

- **Are impacts observable across the board or are they specific to definable subgroups?**

Tables that involve experimental impacts include averages or percentages that are adjusted using linear analysis of covariance procedures controlling for the following seven characteristics before random assignment: race/ethnicity, sex of the focal child, age of the focal child, number of children respondent had given birth to, respondent's initial literacy score, and two sites: Philadelphia and Portland.[13] These baseline characteristics were selected as covariates because they were found to be correlated with observational measures of parenting behavior. Using them as controls in impact analyses helps to statistically equate the experimental and control groups.

In summarizing subgroup impacts in Chapters 4–6, three common distinctions are made: (1) impacts that were significant *across the board*, that is, regardless of the particular subgroups into which particular mothers fell; (2) differences in parenting between experimentals and controls that were detectable primarily among those *comparatively more or less disadvantaged at baseline* (for example, those who had lower or higher educational attainment and literacy skills); and (3) impacts that were significant across a *scattered* set of baseline characteristics that do not represent a discernible pattern.

In Chapter 7, we look across the different parenting measures. Considering those parenting variables for which significant program impacts were detected, we explore the question:

- **How do positive impacts come about?**

In these analyses we ask whether, among families in the experimental group, a particular measure of parenting differs significantly according to extent of participation in particular program components. We then address the question of what unique information has been added by obtaining observational measures of mother-child interaction. Specifically, we ask:

---

[13] Of the seven observational study sites, only the site variables for Philadelphia and Portland were correlated with parenting measures.

- **How closely are observational and interview-based measures of parenting related? Do the observational measures of parenting improve our ability to predict child outcomes beyond the interview-based measures of parenting?**

In these analyses we examine correlations among the differing parenting measures, addressing a sequence of specific questions about their associations. We also carry out multiple regressions predicting selected child outcomes from the 42-month follow-up. In these analyses, we first enter a series of variables to control for child and maternal background characteristics. We then examine the ability to predict child outcomes of the interview-based measures of parenting. As a final step, we add observational measures to the model including interview-based measures of parenting, and examine whether the addition of the observational measures adds significantly to our ability to predict child outcomes above and beyond the interview-based measures.

In Chapter 8 we examine the role played by parenting variables relative to other important factors in predicting the child outcomes. Here we ask:

- **How important are measures of parenting, relative to other important factors, in predicting the development of the children in the sample?**

Again, we rely on multiple regression; however, here we go beyond consideration of parenting variables. In these analyses we consider sets of variables as predictors of selected child outcomes, adding sets of variables cumulatively. We first enter as control variables background characteristics of the mothers and children at baseline. We then add blocks of variables reflecting parenting behavior (both observational and interview-based measures), maternal psychological well-being, maternal human capital, and the broader context of the child and family (for example, residence, social support, difficult life circumstances). The relative merit of each block in predicting the child outcomes is evaluated.

Throughout these analyses we report on the statistical significance of observed relationships or group differences. Statistical significance refers to the notion that an observed relationship between variables, or a difference in the mean score on a variable between two or more groups, is unlikely to have arisen simply by chance. Following common conventions, in this monograph an observed relationship or impact is considered to be statistically significant if there was a smaller than 10 percent probability that the observed relationship or difference could have occurred if only chance were operating.[14]

In Chapter 3 we turn to an examination of the components of the New Chance Program that had the potential to affect mothers' parenting behavior. We focus on, but also go beyond, the parenting education component of the intervention, noting reasons that other aspects of the New Chance Program could have affected parenting behavior.

---

[14] Asterisks are used to indicate levels of statistical significance: one indicates a smaller than or equal to 10 percent probability that the finding arose by chance, two indicate a smaller than or equal to 5 percent probability, and three indicate a smaller than or equal to 1 percent probability. The absence of any asterisks indicates that the observed relationship or difference was not statistically significant. Analyses throughout the monograph that involve tests of significance are two-tailed.

Chapter 3

Participation in Program Components That Could Affect Parenting Behavior

M. Robin Dion, Martha J. Zaslow, and Donna R. Morrison

This chapter presents information for the observational study sample on participation in New Chance Program components. Specifically, we summarize the proportion of mothers who participated in specific program components, the duration of participation, and maternal and family characteristics associated with extent of participation. We confirm that, in general, mothers in the experimental group received services through New Chance significantly more than controls who sought similar services on their own in their communities.

Four hypotheses are articulated for how participation in New Chance Program components could affect parenting behavior. Specifically, parenting impacts may be related to (1) participation in parenting classes alone, (2) participation in both parenting classes and other program components that had content relevant to parenting, (3) participation in program components focusing on human capital development, and (4) total participation across all program components. Experimental group mothers in the observational study spent an average of only 18 hours in parenting education classes, but 143 hours in program components that had some content relevant to parenting, 220 hours in program components focusing on human capital development, and 287 hours across all New Chance Program components. Although the relationship is not strong, it appears that those mothers in the experimental group who were somewhat less disadvantaged at the outset participated in these New Chance activities more.

This chapter also describes in some detail the nature and content of New Chance parenting education classes. Retrospective interviews with field staff were carried out to provide qualitative data on the material covered in parenting education classes, aspects of parenting emphasized, how classes were organized, staffing, whether parenting education was at all linked with child care services, and challenges faced by field staff in implementing parenting education classes. These interviews revealed that developing parenting skills in mothers in the New Chance Intervention proved to be more challenging than had at first been anticipated. According to staff reports, the mothers were cognitively, emotionally, and socially at a considerable disadvantage in terms of being able to accept and integrate the information presented to them.

I. Introduction

Chapter 1 introduced the goals and questions of the New Chance Observational Study within the context of the full New Chance Evaluation, and Chapter 2 focused on the methodology of the observational study. We turn now to an examination of participation in the intervention for participants in the observational study, and especially participation in New Chance Program components that had the potential to affect respondents' parenting behavior.

No intervention can hope to change the behavior of individuals without participation that is of sufficient duration and frequency. Similarly, we must presume that the nature and content of the services received should be related to the outcomes that they are intended to affect. Thus, the purpose of this chapter is to explore both the *quantity* of services that observational study respondents received and the *nature* of services hypothesized to be most closely related to parenting outcomes. Unlike the other chapters of this monograph, information presented in this chapter relies on both quantitative and qualitative data.

Two distinct issues pertaining to frequency, or amount, of participation should be kept in mind. First, the analysis of parenting impacts in New Chance rests on the assumption that experimental group members received significantly more of the services hypothesized to improve parenting behavior than did control group members. It is not a foregone conclusion that this was the case, for several important reasons. Specifically, all applicants to the New Chance Program — prior to random assignment — volunteered for the program, suggesting that mothers in the control group, as well as in the experimental group, were motivated to seek services they needed or desired. Further, control group mothers, while ineligible for services provided by the New Chance Intervention, such as parenting classes and personal counseling, were nonetheless free to seek such services on their own, within their communities. Thus, one issue related to participation frequency involves determining whether experimental group members more often participated at all in different kinds of services than control group members.

A second issue related to participation — equal in importance — involves the average *quantity* of each type of service actually received. That is, what was the dosage of program services among experimentals? Was it sufficiently large to expect differences in behavior? Again, this cannot be assumed to be true, particularly in light of the fact that New Chance was a voluntary rather than a mandatory program. Experimental group mothers, while encouraged to attend, were not required to attend or to fulfill a minimum participation requirement.

Although this monograph focuses on the program's impact on one aspect of expected improvement (parenting behavior), New Chance was a comprehensive intervention involving many different kinds of services expected to help mothers advance toward self-sufficiency and toward better lives for themselves and their children. Which services or combination of services would be expected to affect parenting outcomes? We offer four hypotheses: First, it is reasonable to expect that parenting classes alone could be the single component responsible for influencing parenting behavior. Second, participation in other key components with content relevant to parenting could have added to the effect of parenting classes. Third, program components that were aimed at enhancing the mothers' human capital (for example, employability development and skills training) could affect mothers' interaction with their children. Fourth, parenting behavior could be most closely associated with respondents' total participation in all components of the program.

In the sections that follow, we elaborate upon each of these hypotheses, and then provide a profile of respondents' rates and frequency of participation in the services expected to be related to parenting behavior. Next, the issue of potential self-selection associated with extent of participation in New Chance services is addressed. Finally, a special focus is placed on the nature and content of the New Chance parenting component because, of all the New Chance services that respondents could have received, parenting classes most closely targeted change in parenting behavior outcomes.

II. Four Perspectives Related to Participation and Parenting Outcomes

A. Parenting Education

New Chance parenting classes typically provided information on children's developmental stages, the use of activities and materials to enhance children's cognitive development, and developmentally appropriate guidelines and strategies for shaping child behavior. Parenting classes included both open discussion of issues of concern to mothers and the planned presentation of specific information. Many of the New Chance sites supplemented their classes with opportunities for supervised mother-child interaction so that mothers could receive feedback on their parenting behavior and benefit from the modeling of various approaches by the instructors. Thus, of all the services provided to respondents in the experimental group, parenting classes most directly focused on the improvement of parenting skills and behavior. Clearly, participation in this program component alone can be hypothesized to account for impacts on respondents' parenting behavior. On the basis of this hypothesis, we will examine both frequency of participation in parenting classes and the nature and content of the parenting classes that were offered in the New Chance Program.

B. Parenting Education and Five Other Key Components

New Chance Program components beyond parenting classes included some content relevant to parenting behavior. Specifically, *life skills training* was aimed at improving communication, problem-solving, and decision-making skills in relationships, including the parent-child relationship. *Adult basic education classes*, including pre-GED and GED preparation, focused on improving mothers' educational attainment, but had the potential of affecting literacy behaviors that mothers engaged in with their young children. *Family planning* sessions included discussions of birth spacing as a means of providing each child with sufficient time and attention. Such discussions may have altered mothers' perceptions about the investments they made in their children. Experimental group mothers whose children received *day care* were sometimes required to observe them in a day care center for the purpose of learning about development and appropriate child care strategies, or to interact with their children under the supervision of parenting instructors or day care staff. Finally, *personal counseling* was an important aspect of the program, particularly for those mothers who had emotional problems, such as depression or low self-esteem, or who lacked social support, conditions which have been found to be important predictors of parenting behavior. Improvements in maternal mental health may have facilitated more positive parenting behavior. Mothers could also discuss problems they were facing in their relationships with their children during personal counseling.

Impacts on parenting behavior, then, may be explained not only by participation in the parenting education component of New Chance, but by such participation together with participation in these other components with some content relevant to parenting behavior.

C. Human Capital Development

Researchers have raised the possibility that services aimed at improving mothers' human capital may have unexpected positive effects in areas other than family economic self-sufficiency. Chase-Lansdale, Brooks-Gunn, and Paikoff (1991), for example, suggest that evaluations of human capital development programs for mothers should include assessments of child outcomes, given the possibility that enhanced maternal education, employment skills, and income may affect child development. Although the effects of components such as skills training, work internships, employability development, and basic education have been typically evaluated in terms of economic impacts for families, it is possible that such classes may have unintended positive effects on mothers' parenting behavior. Perhaps by enhancing mothers' interpersonal and cognitive skills, for example, such classes also inadvertently affect the level of social and cognitive stimulation in mother-child interaction. Classes and internships aimed at preparing mothers for employment may also provide feedback to mothers on behaviors that are relevant to gaining and keeping a job, such as restraint in the expression of emotion on the job, self-direction, and mastery of new skills. Mothers may (intentionally or unintentionally) shape their children's behavior in keeping with the behaviors that they themselves are rewarded for in employment development contexts.

D. Total Participation

Yet a fourth possibility exists: that parenting behavior is influenced by the accumulation of the effects of all the program components. That is, while not directly tied to any single aspect of the New Chance Intervention, mothers' parenting behavior may have been affected by the experience of the program *in toto,* perhaps through a sense of overall supportiveness of the program or a perception of personal progress and competence derived from participation in the program. For example, the cumulative effect of contact with peers and encouraging staff members could have led to positive effects on parenting.

III. Frequency and Rates of Participation

A. Participation in Parenting Education

The data in Table 3.1, based on mothers' self-reports regarding program participation, are drawn from the 18-month follow-up interview, but include only the 290 participants of the New Chance Observational Study. As can be seen, the proportion of experimental group mothers who reported participating in parenting classes is far greater than the proportion of control group mothers who sought this service outside New Chance (62.6 and 17.7 percent, respectively).

Table 3.1

Self-Reported Rates of Participation by Experimentals and Controls at 18-Month Follow-Up

	Percent Ever Attended			
Activity	Experimentals	Controls	Difference	p[a]
Components related to parenting				
Parenting classes	62.6	17.7	44.9	0.000 ***
Life skills training	47.2	11.4	35.8	0.000 ***
Family planning	34.9	13.9	21.0	0.000 ***
Personal counseling	32.6	8.4	24.2	0.000 ***
Education[b]	86.7	52.3	34.4	0.000 ***
Day care for the focal child	64.0	30.3	33.7	0.000 ***
Other services				
Skills training	36.8	29.6	7.2	0.220
Unpaid work internship	5.8	3.1	2.7	0.310
Job counseling	47.0	14.6	32.4	0.000 ***
Employability development[c]	---	---		
Health education	44.7	10.1	34.6	0.000 ***
Sample size	184	106		

Source: New Chance 18-month follow-up survey.

Notes: Calculations for this table used data for all 290 respondents for whom there were 18-month follow-up survey data. The sample sizes may fall slightly short of the number reported because of missing or unusable items from some respondents' questionnaires.

For controls, services were obtained at programs or agencies other than New Chance. For experimentals, services were obtained either at New Chance or, if they were served by additional programs, elsewhere.

The percentages are adjusted using linear analyses of covariance procedures controlling for seven kinds of difference in characteristics before random assignment (age of child, maternal literacy, whether mother had more than one child, gender of child, race/ethnicity, Philadelphia site, and Portland site).

Dashes indicate unavailable data.

[a] An F-test was applied to each adjusted difference between average experimental and control group outcomes. The column labeled "p" is the statistical significance level of the difference between experimental and control group outcomes. That is, p is the probability that average outcomes are different only because of random error. Statistical significance levels are indicated as ***<= 1 percent, ** <= 5 percent, *<=10 percent.

[b] Includes adult basic education, GED preparation, high school, college, or other education programs (e.g., proprietary schools).

[c] In New Chance, this service combined career exploration with pre-employment skills training. Respondents were not asked to report on this service at the 18-month interview.

However, the percentage of controls who received parenting instruction within 18 months of random assignment, 17.7 percent, is not trivial. Note also that despite guidelines for attendance, more than one-third of experimentals (37.4 percent) reported never attending parenting classes.

The self-report data do not provide a continuous measure of total number of hours or times attended; thus, we cannot compare an *average* number of hours of instruction or times attended across experimentals and controls. Data providing number of hours of instruction were collected *for experimental group members only,* however, through the New Chance Management Information System (MIS). Before examining those data, it is important to note that there are discrepancies in the percentages of respondents who participated in each of the New Chance Program components between MIS and self-report data. For example, according to self-report, 62.6 percent of experimentals attended parenting classes — roughly 18 percent fewer than the percentage derived from MIS data.[1] MIS data are shown in Tables 3.2 and 3.3.

Table 3.2 indicates that the average experimental group member in the observational study devoted less than 18 hours to parenting education. However, there was a notable range in hours spent in parenting classes among those who received instruction: 19.6 percent of experimental group members did not participate in parenting education at all, 23.9 percent of the young women spent between 1 and 10 hours in parenting classes, 21.7 percent spent between 11 and 20 hours, 13.6 percent spent between 21 and 30 hours, while the remaining 21.2 percent registered more than 30 hours. The MIS data also reveal one outlier case, with 110 hours of parenting

B. Participation in Other Key Components

Table 3.1 also presents rates of participation by experimentals and controls in the other key components that have content directly relevant to parenting behavior (our second hypothesis noted above). It should be noted that experimentals participated in each of the key components (life skills training, basic education, family planning, child care, and personal counseling, in addition to parenting education) significantly more often than controls, who sought these services outside New Chance. There are several other important points to note in this table. First, over half of controls reported having attended some form of basic education classes. For both experimentals and controls, participation in an education program (for example, GED preparation, basic education, high school, or college) was more common than participation in any of the other

---

[1] MIS data were collected by site staff, while self-report data were drawn from the 18-month survey. Self-report and MIS data could differ for several reasons. First, MIS data were recorded concurrently with participation, while self-report data were retrospective, reported 18-months after enrollment in the program; discrepancies could be a result of poor recall. Second, mothers apparently did not categorize activities using the same cognitive scheme as the MIS; for example, they sometimes confused location with activity, as in the case of vocational training that took place at a community college (see Part II for a discussion of mothers' interpretation of survey questions). In addition, MIS data were limited to those activities that were included in New Chance, while the survey data captured participation in other kinds of activities as well. Certain New Chance activities that were reported separately in the MIS data (for example, GED preparation and life skills classes) appear to have been reported as a single activity — education — by some survey respondents, thus inflating the apparent intensity of education services. Finally, the self-report means for participation in Table 3.1 were adjusted for seven kinds of baseline differences, unlike the average rates of participation based on MIS data for experimentals only.

Table 3.2

Participation of New Chance Observational Study Experimentals in Parenting Classes Within 18-Month Follow-Up Period

Activity Measure	Experimentals
Ever participated in parenting education (%)	80.4
Average hours of participation in parenting education	17.7
Percentage distribution of hours in parenting education	
0	19.6
1-10	23.9
11-20	21.7
21-30	13.6
31-40	8.7
41-50	7.6
51-60	2.2
61-110	2.7
Sample size	184

Source: New Chance Management Information System (MIS) data.

Notes: Calculations for this table used data for all 184 experimentals for whom there were 18- month follow-up data, observational study data, and data entered on the New Chance Management Information System, including values of zero for those who were randomly assigned to New Chance, but did not participate.

key services. Nonetheless, the group difference in education participation rates (34.4 percentage points) was statistically significant.

Among all the services considered in the "key components" set, the greatest group differential in participation was observed for parenting classes (44.9 percentage points). Differentials in participation were next largest for life skills, education, and day care (35.8, 34.4, and 33.7 percentage points, respectively). Finally, differences between groups on participation in personal counseling (24.2 percentage points) and family planning (21.0 percentage points) were smallest among the six components thought to have content directly relevant to parenting behavior.

To understand the extent of attendance, or dosage, of experimentals who participated in these key components of the New Chance Program, it is helpful to turn again to the MIS data, the administration records kept by the New Chance operators at each site. Table 3.3 lists average hours of participation in each component, except for hours of day care and personal counseling, which were not recorded by site staff. Clearly, within this set of program components hypothesized to be related to parenting behavior, experimentals spent most of their New Chance participation hours in basic education (100.1 hours). Far fewer hours were spent in the remaining key components for which data in hours are available: 18.1 hours in life skills, 17.7 hours in parenting, as noted above, and 6.7 hours in family planning.

In order to examine participation in day care, we must refer again to self-report data rather than MIS data. Such data were collected for the number of months that respondents received day care services during the follow-up period. Experimental group mothers received day care services for their children for an average of 4.4 months during the follow-up period, a figure that is fairly consistent with the average duration of respondents' participation in the New Chance Program (6.2 months). Control group mothers placed their children in day care settings outside New Chance for an average of 2.4 months.

Although receipt of day care and personal counseling services was not recorded in terms of hours, we might nevertheless create a measure that represents a proxy for average hours in the six key components expected to be related to parenting outcomes. Summing each respondent's MIS hours recorded for parenting, family planning, life skills, and basic education classes, we find that experimentals spent an average of 142.6 hours in the key components of our second hypothesis (excluding children's participation in day care and mothers' personal counseling, which are lacking in the MIS data).

It is interesting to place the data on New Chance day care participation in the context of other major interventions for families in poverty. We have seen that 64 percent of the children of the experimental group participated in day care at some point during the evaluation, but that the duration of care was on average only a little over four months, generally reflecting mothers' limited participation in the New Chance Program. A review of the literature on early childhood programs shows that other interventions have had a much higher level of intensity (Ramey et al., 1995). For example, the Abecedarian Project (Ramey, 1992), the Brookline Early Education Project (Hauser-Cram et al., 1991), and the Infant Health and Development Program (Ramey et al., 1992) all provided care five full days per week, year round, with children participating over a period of years. Unsuccessful early childhood interventions have been associated with a lack of

Table 3.3

Participation of New Chance Observational Study Experimentals in All MIS-Recorded Activities Within 18-Month Follow-Up Period

Activity Measure	Experimentals
Participated in (%)	
Any activity[a]	88.0
Basic education	84.2
Parenting education	80.4
Employability development	79.9
Life skills	78.3
Family planning	75.5
Other group activities	73.4
Health education	64.7
Skills training	39.7
Work internship	22.8
Average hours of participation in:	
All counted activities[a]	286.8
Basic education	100.1
Skills training	56.0
Work internship	38.7
Employability development	24.8
Family planning	6.7
Parenting education	17.7
Health education	7.5
Life skills	18.1
Other group activities	17.1
Percentage distribution of hours in all activities	
0	12.0
1 - 100	24.5
101 - 300	26.1
301 - 500	16.3
501 or more	21.2
Total[b]	100.0
Months in any activity[c]	
Average	6.2
Median	5.0
Sample size	184

Source: New Chance Management Information System (MIS) data.

Notes: Calculations for this table used data for all 184 experimentals for whom there were 18-month follow-up data, observational study data, and data entered on the New Chance Management Information System, including values of zero for those who were randomly assigned to New Chance, but did not participate. Rounding may cause slight discrepancies in sums and differences.

[a] Excludes personal counseling, college classes, and participation in day care.

[b] Distribution does not actually add to 100.0 because of rounding.

[c] Number of months in which experimentals took part in New Chance activities may not have been continuous.

intensity. Furthermore, an analysis of the Brookline project showed that only the most intensive services were found to benefit children who had parents with the lowest educational levels, while a low or intermediate intensity had no measurable consequences (Hauser-Cram et al., 1991).

The quality of on-site child care in New Chance was examined in the larger sample and described in the 18-month interim report. Quality in New Chance preschool classrooms was found to be just below the "good" rating on the Early Childhood Environment Rating Scale (ECERS; Harms and Clifford, 1980) and the Infant and Toddler Environment Rating Scale (ITERS; Harms, Cryer, and Clifford, 1986). The sampled New Chance child care centers generally met or exceeded National Association for the Education of Young Children standards for group size, but fell below desired levels for teachers' educational levels and enrollment-based group size of infant classrooms. Among the seven observational study sites, four were included in this analysis of New Chance day care quality. Two of the seven observational study sites did not provide on-site child care (with the exception of emergency drop-in care), and one site provided on-site child care for only about half of the New Chance children. These sites referred participants to family day care or day care centers or asked them to call an information and referral service to locate care, and many mothers found care on their own. Although New Chance staff were directed to try to guide participants toward developmentally appropriate child care, the quality of care in these settings is unknown.

Thus, both the intensity and quality of on-site New Chance child care appear to have been lower than in previous early childhood interventions that have shown positive impacts.

C. Participation in Human Capital Development Services

We include in the human capital development hypothesis participation in education, as well as training and experiences that were directly aimed at improving mothers' self-sufficiency. In our discussion of rates and frequency of participation in basic education, we noted a large group difference. Table 3.1 indicates that there was no statistically significant difference between experimentals and controls in their participation in skills training or unpaid work internships. However, experimentals attended job counseling classes (47.0 percent) far more often than controls (14.6 percent), a significant difference. Employability development was not measured by self-report.

Turning to the question of extent of participation once again, we note in Table 3.3 that human capital development services were the most frequently attended of all New Chance activities. Average hours of participation were highest in basic education (100.1 hours), followed by skills training (56.0 hours), work internships (38.7 hours), and employability development (24.8 hours). For each of the remaining New Chance Program components, experimental group mothers participated less than 20 hours, on average.

Although a job counseling component was offered but not recorded in the MIS data, a proxy measure for human capital development services was nevertheless created by summing experimental group mothers' hours in basic education, skills training, work internship, and employability development. According to this measure, experimental group mothers engaged in New Chance human capital development activities on average for 219.6 hours.

D. Total Participation

Although we know of no empirical evidence or support in the literature for such a hypothesis, it seems plausible that participation in all the components reviewed above, together with participation in other services not discussed, such as health classes and other group activities, could have created an overall effect that best explains positive impacts on parenting outcomes. For example, most of the activities that come under the human capital development hypothesis relate directly to job preparation. Perhaps the benefits of such services, when added to others that relate more directly to parenting, enhanced respondents' general sense of self-competence or overall confidence, allowing them to pursue parenting in a more positive way. Alternately, the greater overall program contact that comes with participation in all components could have provided needed social support.

MIS data indicate that experimentals in the observational study spent an average total of 286.8 hours across New Chance activities. Again, it should be noted that this figure does not include personal counseling services or day care for respondents' children. The exclusion of contact with day care services is perhaps not serious here, as this would have largely overlapped with participation in other components and would reflect the child's rather than the mother's program contact.

IV. Factors Associated with Extent of Participation by Experimental Group Mothers

As the data above indicate, not all experimentals participated in New Chance parenting education classes or other program components. Furthermore, among experimental group members who did participate, there was considerable variation in the number of parenting education classes and other program classes received. The following section identifies the background characteristics associated with extent of participation by New Chance experimental group members in the observational study, using the four participation perspectives discussed above. Our focus here is on experimental group mothers only and relies on participation as defined using MIS data.

A. Predictors of Participation in Parenting Classes

Using MIS data, New Chance parenting class participation was correlated with a number of background characteristics of respondents obtained at baseline. We report here correlations that are significant ($p < .1$, two-tailed). Participation in parenting classes was found to be significantly positively related to the number of children that the mother had already given birth to at baseline, $r = .14$, $p < .05$. That is, mothers who had more than one child at baseline were also more likely to participate in parenting classes. Mothers who participated more in parenting classes were more likely to have had a focal child who was a daughter rather than a son, as shown by the correlation between gender of child and mother's participation, $r = .13$, $p < .08$. Participation was negatively related to the number of social supports the mother reported that she had at baseline, $r = -.13$, $p < .08$, so that the fewer the sources of emotional support the mother had at baseline, the more she tended to participate in parenting classes. Finally, mothers who had been out of school for two years or less tended to participate in parenting classes more than mothers who had been out of school for three years or more, $r = -.13, p < .07$.

B. Predictors of Participation in the Key Components

As noted above, a proxy measure of participation in the key components hypothesized to have content directly related to parenting was created by adding the number of hours that respondents participated in life skills, family planning, and basic education classes, as well as parenting classes. As above, the proxy was related to a number of baseline background characteristics of respondents. The results showed that, again, participation was significantly negatively related to the length of time that the mother had been out of school, $r = -.13, p < .08$. In addition, the higher the grade the mother had completed in school, the more likely she was to have participated in the key components, as measured by the proxy, $r = .13, p < .09$. Finally, participation was related to history of welfare receipt, such that the less a respondent's family had received welfare as a child, the more likely it was that she participated in the key components, $r = -.14, p < .07$.

C. Predictors of Participation in Human Capital Development Services

As reported above, the proxy measure for New Chance human capital development services excluded job counseling (because hours of participation in this component were not recorded as MIS data) but included basic education, skills training, work internships, and employability development. Participation in these services was associated with mothers' prior employment experience measured at baseline, $r = .17, p < .05$, so that those experimental group mothers who had held at least one job prior to random assignment were more likely to participate in human capital development services. Participation in these services was also related to history of welfare receipt, $r = -.18, p < .05$; the longer the mother's family had received welfare as a child, the less she tended to participate in human capital development services.

D. Predictors of Total Participation

As indicated earlier, a measure for total participation was created by summing participants' hours in all New Chance activities, except for components for which MIS data were unavailable. Total participation was found to be most highly related to whether or not the participant had ever been employed at baseline, $r = .18, p < .05$. Mothers were more likely to participate in New Chance activities overall if they had been employed at some point prior to random assignment. Mothers' participation in New Chance was associated with less welfare receipt in their family histories, $r = -.17, p < .05$. Finally, the relationship between total participation and completion of at least the 11th grade in school at baseline approached significance, $r = .12, p < .11$.

V. The Nature and Content of the New Chance Parenting Education Component

Because parenting education was the New Chance Program component that most directly and intentionally targeted the outcomes of interest to us, we turn now to a more detailed description of this program component.[2] We seek here to portray field staff accounts of the goals and challenges involved in providing New Chance mothers education in parenting.

A. Content

The primary goal of the parenting education component of the New Chance Program was the improvement of parenting by enhancing respondents' ability to foster their children's cognitive, social, emotional, and physical development. The general nature of the parenting classes has been described in prior MDRC publications, namely, *New Chance: Implementing a Comprehensive Program for Disadvantaged Young Mothers and Their Children* and *New Chance: Interim Findings on a Comprehensive Program for Disadvantaged Young Mothers and Their Children.* The following section includes information from those reports and adds further detail obtained from a recent set of structured telephone interviews with field staff at the observational study sites.[3]

MDRC trained parenting instructors with a curriculum designed for use with disadvantaged young mothers: *A Guide to Helping Young People Parent*, developed by the New York City Department of Health, Bureau of Maternity Services and Family Planning. According to the field staff interviews, all observational sites based their parenting component on this curriculum, which included units on child development (for example, age-appropriate behavior, children's feelings and self-esteem, how children learn), developing values in children, preventing child abuse and the role of discipline, nonsexist childrearing, information about children's sexuality, and addressing children's medical needs and dealing with accidents and emergencies. There were also tips concerning toys and play activities. Many of the observational study sites supplemented this core curriculum with other materials, including *Systematic Training for Effective Parenting* (STEP), *Parenting Black Families*, *The Nurturing Parent Program for Teen Parents*, *SOS! Help for Parents*, and *Without Spanking or Spoiling*.

The guidelines for parenting classes discouraged reliance on a lecture format for instruction and encouraged active, participatory sessions. According to field staff interviews, most parenting instructors found that providing structure facilitated the mothers' learning process. Thus, they typically sought to achieve a balance between structure and participatory interaction. All observational study sites contacted for interviews reported that they followed a planned outline for parenting class sessions, but remained open to addressing issues that the mothers brought up themselves.

---

[2] Of course, we have no way of assessing the content or quality of parenting instruction that controls may have received outside the New Chance Program.

[3] A number of factors should be kept in mind when evaluating information obtained in the field staff interviews. First, the interviews were qualitative in nature. Respondents were asked a standard set of open-ended questions that were aimed at eliciting information about each site's unique implementation of the parenting component. Second, because the interviews were conducted more than three years after initiation of the demonstration, they should be considered retrospective. Third, since it was not always possible to interview the parenting instructor and/or day care director directly, site directors sometimes served as respondents, resulting in some variability in respondents' level of information. Fourth, the set of questions asked of respondents was not comprehensive, but rather was selective in an attempt to inform the findings of the observational study. Finally, staff at only six of the seven observational study sites could be contacted for interviews.

B. Challenges

Interviews with field staff revealed that most of the parenting instructors found it very challenging to teach this population of young women for several reasons. First, instructors indicated that the young mothers often had trouble processing information and seemed to need much more time than was initially allotted to understand and integrate the material. Second, the mothers were typically mistrustful of the parenting instructor at first. It usually took some time and concerted effort for instructors to establish rapport with the young women. Third, the mothers often appeared to be experiencing emotional problems, such as low self-esteem, isolation, loneliness, and depression. One site director summed it up by stating that a young woman's progress in the overall program depended almost entirely on the state of her mental health. If she was feeling very badly about herself, she experienced great difficulty in resolving any of her problems — whether it was a housing problem, boyfriend trouble, or poor parenting skills. According to this director, a mother's ability to move forward in the program often depended on her ability to address such problems.

A fourth issue regarding parenting that was raised repeatedly by field staff at the observational study sites was families' lack of support, or actual interference. Specifically, in trying to apply their new parenting skills, the young women were often undermined by their own mothers or experienced conflict with them. Grandmothers were often unfamiliar with such child discipline techniques as "time-out" and sometimes pressured the young women to use methods that they themselves had been raised with — methods that were sometimes harsh, abusive, or overly permissive. The problem appeared to extend to others in the family. One young woman, for example, commented to a field supervisor: "I don't want to hit my kid, but other people in my family want me to."

Finally, participants' fundamental understanding of children's development at the outset appeared to be fairly poor. For example, some sites found it difficult to teach appropriate behavior management until emotional development had been addressed. One mother thought that a baby who kicks during diapering is angry and should be disciplined. Another mother had difficulty understanding why her child would be afraid of a Halloween mask when it didn't seem scary to her. Other sites found it difficult to teach cognitive stimulation skills before parents understood how to manage their children's behavior. For example, when an instructor tried to teach parents how to read to their children, mothers wanted to know: "But how do you get him to sit still?" Indeed, several sites indicated that teaching appropriate child discipline tended to take the most time and energy to cover, because it usually engendered much controversy among the mothers.

C. How Sites Addressed the Challenges

Although sites found parenting education to be more challenging than anticipated, most sites modified their approach to address these challenges. For example, at least one site revised the sequence of learning modules so that emotional development was presented first, followed by behavior management, and finally cognitive development and other issues. Site directors learned to choose parenting instructors with care. Directors did not find, in general, that age, race, or even educational credentials[4] of the instructor were as important to an instructor's effectiveness as her basic temperament. They felt it was critical that the individual be one of "strong character," so that she could facilitate discussion and yet maintain control. One issue that related to the mothers' trust of the instructor was whether or not the instructor had children of her own. The mothers tended to be suspicious of instructors who were not parents themselves. This issue was addressed in various ways. At one site, the instructor acknowledged from the start that she could never know what it is like to raise children of one's own. Another teacher emphasized that she had worked with New Chance children all day, every day for long periods, and thus had much experience learning from other mothers from which she could draw.

One site arranged for mothers to be clinically tested for depression, and mothers who were found to have high levels of depression were referred out of the program for treatment. Another site attempted to gain the support of the mothers' boyfriends or partners.

After the end of the demonstration phase, many of the sites revised their parenting component to reflect what they had learned. This included breaking the units into smaller steps and allotting more time per unit. Some sites added new units and topics based on issues that mothers repeatedly raised (for example, nutrition, child safety, being a good parenting consumer). Others made changes in an attempt to involve the young mothers' families and added "individual parenting sessions." One site director added modules that related to personal development — that is, assertiveness training, anger management, alcohol abuse, and women's studies — and remarked that one result has been greatly improved attendance.[5] Finally, the sites contacted for interviews often concluded that given the challenges associated with teaching this population of mothers, much more time than the suggested five months was necessary to cover the material effectively.

D. Mother-Child Interaction

Four of the six observational study sites where interviews with staff were possible indicated that they provided regular opportunities for parents and children to interact in the presence of staff members who could provide feedback. The form of these opportunities varied. At one site, mothers were required to have breakfast together with their children each morning. At another, "parenting labs" were held in which children were brought into the parenting classes twice a month and engaged with their mothers in a variety of structured activities, such as nature walks, sensory stimulation for infants, and games with toddlers. At other sites, parents were asked to observe or interact with their children in the on-site day care center; for example, parents observed their children at play through a one-way window. At one site, mothers taught their children to play jump rope, practiced reading to them, and worked with them in planting a vegetable garden. This same site was successful in integrating mother-child interaction with the participants' basic education and parenting education. For instance, mothers were asked to research dinosaurs at the library as part of their own education, and then were required to write a report in language that a very young child would understand. Next, mothers completed a parenting unit on

---

[4] Credentials of parenting instructors typically included college degrees in psychology, education, or home economics. Many had experience teaching inner-city young people.

[5] These modules may have been added as a result of the life skills component being discontinued after the demonstration phase.

"how to take your child to the museum," and subsequently participated in a field trip to a museum of natural history with their children where they both learned more about dinosaurs.

Four of the six sites that were interviewed had on-site day care during the 18-month follow-up period. The remaining two sites had only limited day care available on their premises. These sites found it difficult to schedule opportunities for mother-child interaction beyond accompanying children on field trips or special events, such as a picnic or cookout.

E. Class Structure and Requirements for Length of Attendance

There was variation across sites in length of a typical class session and frequency of classes. Classes usually lasted for 1½ to 2 hours and met weekly or twice weekly. Average class size ranged from about 15 to 30 participants, but often varied significantly because of attendance problems. For example, two sites encouraged mothers to participate in parenting classes for five months, while another site considered parenting classes to be "open entry and open exit," that is, requiring no fixed number of parenting classes or hours. Yet another site required mothers to attend parenting classes as long as they were taking GED classes.

In sum, information from the field staff interviews suggests that certain aspects of the parenting education component varied somewhat across the seven observational study sites.[6]

VI. Summary and Conclusions

In this chapter, we have examined participation from four distinct perspectives. We have hypothesized that parenting impacts may be related to participation specifically in parenting classes, to participation in parenting classes and other key components with content relevant to parenting behavior, to participation in human capital development services, or to total participation in all program components. The participation findings for the observational study sample indicate that experimental group mothers received most services significantly more often than controls who sought similar services on their own in their communities. Yet we also find that hours of participation in parenting education were fairly limited, particularly compared with the hours of participation in basic education or job-related services that respondents received. Among experimentals, dosage varied with each of the participation hypotheses: experimentals spent an average of just 17.7 hours in parenting classes, 142.6 hours in the key components (parenting education, life skills training, family planning, and basic education), 219.6 hours in human capital development services (basic education, skills training, work internships, and employability development), and 286.8 hours in the total program (excluding day care and personal counseling).

We have also reviewed information concerning the content and nature of New Chance parenting classes. In general, the field staff interviews revealed that teaching parenting skills to this population of young women proved to be more challenging than had at first been anticipated. For cognitive, emotional, and social reasons, the mothers were at a considerable disadvantage in terms of being prepared to accept and integrate the information presented to them.

---

[6] Such variation could have resulted in slightly different outcomes across sites. Analyses that examine impacts for individual sites could be undertaken. However, given that the relatively small sample size of observational study participants at each site would limit the ability to detect statistically significant findings, we have not reported such analyses here.

Indeed, it appears that a combination of several factors should temper the expectation that positive impacts on parenting behavior resulted directly or solely from participation in parenting classes. Specifically, a substantial number of mothers in the experimental group did not receive any parenting classes at all (19.6 percent according to MIS data, or 37.4 percent according to self-report); many control group mothers did receive parenting classes in the community (nearly 17.7 percent according to self-report); the average number of hours attended, among those who participated at all, was relatively low, and the population faced many challenges in responding to parenting classes.

We have seen that among the mothers in the experimental group, certain background characteristics were related to extent of participation. Although the relationships were generally not strong, it appears that those experimentals who were somewhat less disadvantaged at the outset participated in New Chance activities more often. Mothers who participated more tended to have completed a higher grade in school, to not have received welfare as a child, to have been employed at some point prior to random assignment, and to have been out of school for two years or less at baseline.

Later in this monograph, we will use the four perspectives related to participation in an attempt to understand the specific mechanisms underlying any program effects on parenting behavior, and analyses will adjust for the selection factors noted in this chapter (see Chapter 7).

Chapter 4

The Affective Quality of Mother-Child Interaction

Nancy S. Weinfield, Byron Egeland, and John R. Ogawa

This chapter examines the affective quality of mother-child interaction during an observation carried out approximately 21 months after random assignment. Findings are presented in several ways. First, we look at how selected characteristics of the New Chance and control group participants relate to affective quality. Second, we examine the impact of assignment to the New Chance Program on affective quality. Finally, we explore impacts within selected subgroups of participants to see whether assignment to the program has affected subgroups differentially. Results indicate that those participants assigned to the program were less harsh with their children during the observation.

The first three chapters in this monograph provided an introduction to the New Chance Observational Study, introducing its goals, presenting the methodology of the study and demographic characteristics of the participants, and examining differences within the observational study sample between experimental group and control group members in program participation. This chapter examines data on the affective quality of the mother-child relationship at approximately 21 months after random assignment.

I. The Importance of Studying the Affective Quality of Mother-Child Interaction

A. Background

Psychologists have long believed that a child's adaptation to the challenges of development is determined in part by qualities of the early parent-child relationship. Affective warmth, sensitivity, guidance, and structure by parents have been found to be among the most important aspects of parent-child interaction in determining a child's later competence in the face of new challenges (Bowlby, 1969, 1982; Baumrind, 1989). Research and theory in this area draw from two major perspectives: parenting behavior (Baumrind, 1971) and parent-child attachment (Bowlby, 1969, 1982).

Research on parenting behavior has demonstrated that the children who interact the most competently with their peers and their environment come from homes where the parents show both high warmth toward their children and high expectations for maturity and socialization. Diana Baumrind's research (1971, 1989) on parents and their preschool-aged children showed that parents who were both affectively warm and placed high demands on their children for appropriate behavior had children who were self-reliant, self-controlled, content, and open to exploration. Parents who placed high demands on their children yet demonstrated low warmth and sensitivity had children who were disconnected, withdrawn, distrustful, unhappy, and unfriendly. Parents who showed high warmth but placed low demands on their children's behavior had children who were immature, low on self-reliance and exploration, and low on self-control. In addition, Baumrind (1989) found that coercive behavior on the part of the parent was one of the most detrimental types of parent-child interaction. According to Baumrind, coercive behavior undermines the child's internalization of socialization norms because it focuses the child's attention on the power of the parent rather than the importance of the socialization issue at hand.

Attachment theory focuses on the patterns of interaction between parents and their young children as laying the groundwork for children's expectations and behaviors in subsequent relationships and social interactions. Attachment theory holds that sensitive, responsive interactions between the parent and infant over time will lead the infant to expect that others are sensitive to his or her needs and cues and that he or she is deserving of such consideration. This confidence in the availability of the parent (or caregiver) allows the child to use her as a base from which to explore the world, secure in the expectation that should danger arise the child will be protected and comforted by the parent (Bowlby, 1969/1982). The quality of early attachment is described as secure or anxious, with the anxious classification further divided into anxious/avoidant and anxious/resistant.[1] From security in this early relationship the child gains a sense of efficacy in approaching the world, and acquires the confidence and curiosity to take on new challenges. As the child moves beyond infancy, Bowlby suggests that the attachment relationship changes in its manifestations and balance. He suggests that the relationship moves into a goal-corrected partnership, where the secure child is able to consider the mother's goals and activities when making bids for nearness or attention, knowing from experience that his or her needs will be met as soon as possible. As a result, the child can delay gratification briefly to account for the mother's goals. The dyad is, in effect, working together to balance the child's needs with practical constraints on the mother's availability. The quality of early attachment relationships is a significant predictor of social developmental outcomes in children (Sroufe, 1988).

Working from these theoretical backgrounds, we are interested in many dimensions of the mother-child relationship and how the mother and child adapt to each other's needs and demands in order to achieve their goals. The affective quality coding scheme used in the New Chance Observational Study focuses on this partnership between the mother and child. From observing the mother's behavior, we are interested in how she guides the child through the teaching session, what kinds of support she gives, whether she encourages autonomy, and whether hostile or coercive techniques are used. From observing the child's behavior, we hope to tap the child's ability to comply with the mother's socialization demands, exercising the persistence, self-control, and affect that have been internalized from the ongoing mother-child relationship. From observing dyadic behavior, we seek to explore how the dyad negotiates challenges together, whether through synchronous interaction, conflict, or disengagement.

---

[1] In attachment relationships described as "secure" the caregiver has been available and comforting when the infant needs comfort. Consequently, the infant will continue to seek the caregiver when distressed and will be calmed by contact with the caregiver. In attachment relationships described as "anxious/avoidant" the caregiver has been chronically rejecting or emotionally unavailable to the infant's cues. Consequently, the infant responds to being distressed by not seeking contact with the caregiver, and even avoiding contact or proximity if it is offered. In attachment relationships described as "anxious/resistant" the caregiver has been inconsistent in responding to the infant's cues. Consequently, the infant may be unduly vigilant with regard to the caregiver's whereabouts. When distressed the infant is not easily comforted and may alternate between seeking contact with the caregiver and rejecting such contact angrily.

Rating scales examining mother-child interaction on the Teaching Tasks were developed for use on the Minnesota Mother-Child Project, a longitudinal study of high-risk mothers and children (Egeland and Brunnquell, 1979). The mothers and children of the Mother-Child Project were observed in the Teaching Task situation when the children were 42 months old, and the sessions were coded for many of the dimensions of maternal behavior, child behavior, and dyadic interaction mentioned above. (See Table 4.1 for brief descriptions of the rating scales as adapted for the New Chance Evaluation.) There has been much research done on the Minnesota Mother-Child Project with these ratings. This research has used Teaching Task ratings as correlates, to illuminate group differences, as intermediate variables, and as predictor variables. This research provides broad evidence for the validity of the Teaching Task ratings in research, and their relevance for predicting policy-relevant outcomes. In order to provide a context for understanding the measures of affective quality in the New Chance Observational Study, we will summarize briefly the findings from previous research using the Teaching Task ratings.

B. Predicting Affective Quality on the Teaching Tasks: Correlates and Group Differences

Rahe (1984) found that the quality of mother-infant attachment, assessed during infancy, was related to a number of the Teaching Task variables. Children who were anxious/resistant in their attachment to their mothers in infancy were rated as less persistent and compliant and more dependent on their mothers during the Teaching Tasks than secure or anxious/avoidant children. Anxious/resistant children were also less enthusiastic than children who had secure attachment histories, while those with anxious/avoidant histories fell between the other two groups on the Enthusiasm scale. Rahe also found that the Teaching Task performance was predicted by several of the scales on the Home Observation for Measurement of the Environment (HOME) Inventory (Caldwell and Bradley, 1978), a measure of the environment and interactions in the home setting.

Farber and Egeland (1987) examined differences on the Teaching Tasks between children who had been abused by their parents and a non-abused matched control group. They found that, in comparison with the control group, abused children were more negativistic, less compliant, less affectionate toward their mothers during the session, more avoidant of interaction with the mother, and less persistent and enthusiastic in engaging the tasks. Their findings differentiated the groups on an even finer level, showing that within the abused group children who had the buffer of a secure early attachment relationship appeared more competent on the Teaching Tasks than those with an early insecure attachment.

C. Affective Quality on the Teaching Tasks as an Intermediate Variable

Erickson, Sroufe, and Egeland (1985) predicted preschool behavior problems from infant attachment classifications. They found that those children who defied prediction (that is, secure children who exhibited behavior problems in preschool or insecure children who did not exhibit behavior problems in preschool) showed patterns of behavior on the Teaching Tasks (at 42 months) that were significantly different from the other children who shared the same early attachment classification. Secure children who exhibited behavior problems in preschool were less affectionate and more avoidant and had a poorer overall experience of the session, while their mothers provided poorer quality of instruction, set poorer limits, and were less confident than those dyads with secure children who showed no behavior problems in school. Insecure children who showed no behavior problems in school were more persistent, compliant, and affectionate and had a better overall experience of the session; while their mothers were more respecting of their autonomy, offered more support, set clearer limits, showed less hostility, more confidence, and better quality of instruction than those dyads with insecure children who did show later behavior problems in school. Thus, the Teaching Task ratings seem to be a useful indicator of intermediate change in the relationship.

Egeland and Kreutzer (1991) used the ratings as a part of a composited early competence variable. They found that the summary rating of high early competence predicted the effects of high maternal stress in the early school years. High early competence acts as a buffer against the effects of high maternal stress. Children who were highly competent on the Teaching Tasks experienced fewer of the long term effects of high maternal stress than those who were less competent. The qualities of the relationship captured through the Teaching Tasks seem to act as a protective factor against the effects of difficulties in the mother's life.

D. Affective Quality on the Teaching Tasks as a Predictor Variable

The Teaching Tasks have been used in research to predict several school outcomes, both academic and social, that may be particularly relevant to policy issues. Pianta et al. (1990) used the Teaching Tasks to see if they could differentiate which children would later be referred for special services in the early school years. They found that in comparison to dyads where the child was not referred, children who were later referred for special services were less persistent, compliant, and affectionate, more negativistic and avoidant of interaction with their mothers, and had a poorer overall experience of the session, while their mothers offered less support, showed less respect for the child's autonomy, offered less structure and poorer limit setting, were more hostile, showed poorer quality of instruction, and had less confidence. Thus, qualities that are captured in the Teaching Tasks have implications for later problems in the school setting.

The Teaching Tasks scales have also been used in composite variables to represent elements of children's early experience. Renken et al. (1989) used the ratings from the Teaching Tasks as elements of composite variables of children's early experience while predicting behavior problems in school during middle childhood. They found that the composite variables explained a significant portion of the variance in aggression for boys and for girls during middle childhood and also explained a significant portion of the variance in passive withdrawal for boys.

Pianta and Egeland (1994) examined predictors of changes in IQ from the expected continuity between measures of IQ from 24 to 48 months and from 48 to 96 months. They composited all of the mother ratings from the Teaching Tasks to form a rating of mother's interactive competence with her child during the session. They found that across both time periods, the Teaching Task composite variable accounted for a significant portion of the variance in predicting changes in IQ score, with high interactive competence predicting increases in IQ.

E. Consideration of the Affective Quality of Mother-Child Interaction in the Context of the New Chance Evaluation

By providing a comprehensive intervention that included education, job and life skills training, and parenting classes, the New Chance Program sought to help mothers gain the economic and social resources and skills necessary to help better their lives and the lives of their children. New Chance also sought to benefit the lives of the children more directly. Its two-generational approach included either providing or helping the mothers locate good-quality day care for their children, so that the children would have the added benefit of a day care setting that fostered their development.

McLoyd (1990), in her writings on mothers in poverty, suggests that poverty and the mother's emotional state combine to affect mother-child interactions and negatively influence child development. We expect that a program such as New Chance should reduce these negative effects on mother-child relationships. These changes may come about in two ways: (1) by teaching mothers new academic, job, and personal skills, New Chance should reduce maternal stress and poverty and increase the mothers' personal resources, allowing them to focus more on their relationships with their children; (2) through New Chance, mothers should learn how to relate more positively to their children, even if their lives continue to be somewhat difficult or stressful. As mentioned above, Egeland and Kreutzer (1991) found that the Teaching Tasks are a useful indicator of relationship qualities that protect the child from the effects of maternal stress. It should be informative to use the Teaching Tasks affective quality coding scheme as an indicator of whether the support and skills gained by mothers enrolled in New Chance allow them to improve their relationships with their children.

In the next section we describe the Teaching Task session, the coding scheme used to evaluate the sessions, and the coding procedures followed. Then we present results of analyses on the affective quality variables, as well as detailed analyses on a subset of those variables. We present means on the affective quality variables by baseline subgroups and by 18-month follow-up subgroups. We also present program impacts comparing the New Chance experimental and control groups overall, as well as program impacts found between and within baseline subgroups.

Overall, we found that some maternal characteristics at baseline and 18 months predicted more positive interactions during the Teaching Task session. In addition, we found that mothers who were assigned to the experimental group within the New Chance Program treated their children less harshly than mothers who were assigned to the control group.

II. Procedures

A. The Teaching Task Session

Affective quality of the mother-child relationship was coded from videotapes of the Teaching Task observations. The Teaching Tasks were designed and used originally in research on parent-child interaction by Harrington, Block, and Block (1978). The somewhat modified Teaching Tasks for this observation consist of a book reading task, followed by four teaching tasks and the presentation of a gift. The tasks themselves involve several different activities, such as getting the child to build blocks that resemble a model block out of a variety of smaller pieces; getting the child to name things that have wheels; having the child match plastic pieces to a guide by color, shape, and size; and having the child use an Etch-a-Sketch to draw a line through a maze (for more details, see Chapter 2 in this monograph). The Teaching Task session is a mildly to moderately stressful situation, as it requires the mother to structure the situation so that her child can understand and achieve the set goals of the tasks. This format allows us to see how the mother balances the emotional needs within the relationship with the practical constraint of having goals to achieve. The need for such a balance is common in daily life, thus the Teaching Tasks may reveal what strategies the mother resorts to in the face of such challenges.

B. Coding Procedures

All coders underwent extensive training involving viewing and discussing many tapes, as well as preliminary coding to establish initial reliability. Subsequently, each taped session was viewed and coded independently by two coders who were blind to group membership. For the affective quality coding scheme, coders watched the entire session on videotape several times before assigning ratings, which were based on the events of the entire session. Coding pairs were assigned on a rotating basis, ensuring that all possible coding pairs within the group were represented. Each week coding pairs reviewed the scores that each coder had assigned to a tape; any disagreements were resolved through discussion, review of notes from the tape, or watching actual segments of the tape in question. Through this conferencing process the coding pairs arrived at one score for each scale that they believed best represented the events of the Teaching Task session.

Disagreements between the two coders that took the form of either highly discrepant scores on a tape (three points or more on a 7-point scale) or inability to reach a consensus were resolved by a third coder. In such cases a third member of the coding group viewed and coded the tape in question without having seen the scores of either of the original two coders. The three coders then met to discuss the tape and reach a consensus regarding the scores. Although this process was not often necessary, it proved to be an extremely effective method for dealing with the coding of highly ambiguous or difficult tapes.

In addition to the pair coding, each week one tape was coded by the entire group. For this tape, each member of the group coded the tape independently, and consensus was reached in a manner similar to that used with pair tapes. The individual scores were recorded, and through scale-by-scale discussion the group arrived at one set of scores they believed best represented the events of the Teaching Task session. These group tapes served as ongoing training and to prevent coding drift within the group. They also served as a forum for discussion of issues pertaining to the scales and the coding of particularly ambiguous situations.

C. Rating Scales

Sessions were coded on 14 rating scales (13 7-point scales and one 3-point scale) and a multiple-item checklist of low frequency (often negative) behaviors. While the scales were adapted from the original scales designed for use on the Minnesota Mother-Child Project, some scales were added or deleted to reflect the goals of the New Chance Evaluation. (See Table 4.1.) For example, a scale measuring the child's reliance on the mother's help during the session was eliminated because the broader age range of the children in the New Chance sample made decisions about age-appropriate help-seeking behavior more individual to each case and less well suited for comparison across pairs. The Quality of Relationship scale, which measures behavioral and affective sharing and synchrony between the mothers and children, was added to extend the examination of affective interactions. The checklist items[2] evaluated the occurrence or non-occurrence of low-frequency behaviors such as highly punitive behavior by the mother or signs of psychopathology in the mother or child.

D. Choosing a Subset of Variables for Detailed Analyses

The analytic plan for this chapter (and the other results chapters) includes multiple analyses using each of the rating scales as dependent variables. In order to limit the possibility of chance findings and to consolidate the information, we decided to present detailed analyses only on a subset of the scales. Five scales were chosen that reflect the main goals of the evaluation.

In selecting the five variables, we chose two maternal variables, two child variables, and one dyadic variable that we felt best represented important features of the session. The most salient maternal behaviors we wished to capture were maternal warmth and emotional support, and maternal punitive and coercive behavior. To represent maternal warmth, we chose to do detailed analyses on the Mother's Supportive Presence scale. In order to represent maternal punitive and coercive behavior, we combined scores from the Mother's Hostility scale and six items from the checklist to form a new scale: Mother's Harsh Treatment. The Mother's Harsh Treatment score was derived by adding one point for each of the checklist items that occurred during the session as well as adding one additional point if the Mother's Hostility score was greater than three, resulting in a new scale with a possible range from zero to seven.

The most salient child behaviors we wished to capture were those that related to child socialization and the emotional tone of the child's behavior toward the mother. Consequently, we chose the Child's Compliance scale and Child's Affection to Mother scale for detailed analyses. Finally, to capture a dyadic element of the affective and behavioral qualities of the relationship, we chose the Quality of Relationship scale.

To ensure that the subset of variables chosen was appropriate statistically as well as

---

[2] For the purposes of this chapter, only the subset of checklist items relating to punitive behavior will be addressed: mother taunts child; mother physically punishes child; mother handles child roughly; mother verbally abuses child; mother threatens child; mother nags child.

Table 4.1

Teaching Task Rating Scales

Scales Evaluating Mother Behavior	**Mother's Supportive Presence:** assesses the degree to which the mother expresses positive regard and emotional support to her child. She may do this by acknowledging the child's accomplishments on tasks or on unrelated activities, encouraging the child with positive emotional expressions, or various other ways of letting the child know that he/she has her support and confidence to do well in the setting. **Mother's Intrusiveness:** assesses whether the mother lacks respect for the child as an individual and fails to understand and recognize the child's effort to gain autonomy and self-awareness. There are many ways in which a mother may intrude. For example, intrusiveness can occur through harsh physical treatment, with untimely affection, or if the mother does not allow the child autonomy in the problem-solving tasks. **Mother's Hostility:** reflects the mother's expression of anger, discounting or rejecting of the child. Hostility may take the form of overt rejection of the child, blaming him or her for mistakes, or the mother otherwise making it explicit that she does not support the child emotionally. **Mother's Quality of Instruction:** evaluates how well the mother structures the situation so that the child knows what the task objectives are and receives hints or corrections while solving the problems that are: (a) timely to his/her current focus, (b) paced at a rate that allows comprehension and use of each hint, (c) graded in logical steps that the child can understand, and (d) stated clearly without unnecessary digressions to unrelated phenomena or aspects of the task that might only confuse the child. **Mother's Confidence:** assesses the degree to which the mother seems to believe that she can work successfully with the child in the situation and that the child will behave appropriately (whether this is more or less task-oriented depends on mother's definition of the situation as a social- or achievement-oriented activity).
Scales Evaluating Child Behavior	**Child's Persistence:** measures the extent to which the child is actually problem-oriented in the session. The child may be either sober or playful, compliant or not compliant to the mother's directions, as long as he or she shows motivation toward completing the tasks. **Child's Enthusiasm:** reflects the degree to which the child acts with vigor, confidence, and eagerness to do the tasks, taking an active interest in his/her activities and investing effort in them, as well as appreciating successes. Enthusiasm involves both a sense of agency and coordination between affect and behavior. **Child's Negativity:** assesses the degree to which the child shows anger, dislike, or hostility toward the mother. This may take the form of forceful rejection of the mother's ideas, showing angry and resistant expressions, pouting, whining, or being unreasonably demanding or critical of her. **Child's Compliance:** measures the degree to which the child shows willingness to listen to the mother's suggestions in the setting and to comply with her requests in a reasonable manner. The compliant child is attentive to mother and structures his or her activity around the mother's directions

(continued)

Table 4.1 (continued)

	Child's Experience of Session: reflects the degree to which the child's experience in the session probably resulted in feelings of success and competence on the tasks and confidence in having a good relationship with his or her mother. Both the child's behavior and the mother's behavior toward the child contribute to the child's experience of the session. **Child's Affection To Mother:** reflects whether the child displayed a substantial period of positive regard and sharing of happy feelings toward the mother. Behaviors considered for this scale are overtures and attempts by the child to share positive affect with the mother, such as looking at mother, making eye contact, smiling, and other affective "approach" behavior. **Child's Avoidance of Mother:** reflects the child's tendencies or clear attempts to avoid interacting with the mother in the session. Signs of avoidance include the child showing a tendency, at some point in the session, to withdraw from the mother either by leaving the situation or resisting the mother's attempts to engage him or her.
Scales Evaluating Dyadic Behavior	**Quality of Relationship:** focuses on affective and behavioral reciprocity of the mother-child relationship. High quality of relationship is indicated by a strong sense of relatedness and mutual engagement between mother and child, with each explicitly acknowledging and responding to the other. Any conflicts are quickly, easily, and amicably resolved with little or no escalation. **Physical and/or Psychological Dissolution of Boundaries in the Parent-Child Relationship:** evaluates the degree to which the parent and child maintain appropriate role relationships. There are two forms of boundary dissolution, psychological and physical. The psychological boundaries between a mother and her child may dissolve when the mother begins treating the child as her contemporary (either playmate or intimate partner) rather than taking charge and setting the necessary limits. The physical boundaries between a mother and her child may dissolve when the mother controls or manipulates her child using physical intimacy and sensuality.

theoretically, we inter-correlated all the scales. This step allowed us to examine how well these five variables represented the information available from all the scales. The correlations are displayed in Table 4.2. The five scales chosen for detailed analyses: Mother's Supportive Presence, Mother's Harsh Treatment, Child's Compliance, Child's Affection to Mother, and Quality of Relationship were highly correlated with the remaining scales. These five variables captured most of the information from all of the affective quality scales and therefore could be used to limit the number of analyses to be presented. The choice of these dimensions of maternal, child, and relationship qualities were further confirmed through factor analysis, although the affective quality scales were chosen for analyses over the factor scores to preserve the clarity of interpretation.

E. Reliability

Interrater reliability (see Table 4.3) was determined using the original scores assigned by the two coders of each tape. The statistic used to determine reliability was intraclass correlation. Reliability on interval rating scales is best evaluated using intraclass correlation, as statistics such as Kappa are intended primarily for use with ordinal and nominal data. The Mother's Confidence scale, unlike the other scales, is a 3-point scale. As a result, intraclass correlation is not the best way to represent reliability for this scale. Instead, reliability for this scale is represented with Kappa. Reliability for each scale was within acceptable ranges.

III. Findings and Discussion

Analysis of the scores from the affective quality scales proceeded in four stages: (1) computing descriptive statistics for each of the scales, (2) examining how different subgroups (split by demographic and psychosocial variables measured at baseline and the 18-month follow-up) score on the affective quality scales, (3) testing whether the New Chance Program had an effect on the affective quality of mother-child interactions, and (4) testing whether the New Chance Program had differential effects on the affective quality of mother-child interactions for different baseline subgroups.

A. Descriptive Statistics for the Affective Quality Scales

Descriptive statistics (mean, standard deviation, and potential range) for each of the 15 affective quality scales are presented in Table 4.4. Data for all the rating scales are presented in this table; in subsequent sections we focus on the five scales chosen for more detailed analysis. Most scales (Mother's Supportive Presence, Intrusiveness, Quality of Instruction, and Confidence; Child's Persistence, Enthusiasm, Compliance, Experience of Session, and Affection to Mother; Quality of Relationship) had an approximately normal, bell-curve distribution. Mother's Hostility, Mother's Harsh Treatment, Child's Negativity, Child's Avoidance of Mother, and Boundary Dissolution were distributed approximately quadratically, with a large number of participants scored at the lower scale points and a curve leading to only a few participants who were scored at the higher scale points. The quadratically distributed scales rate primarily negative aspects of the mother-child relationship. Thus, the ratings of positive and negative aspects of the interactions are distributed differently in the sample, a finding that parallels research from other samples.

Table 4.2

Interscale Correlations for Affective Quality Scales

Observational Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Ratings of the Mother															
1 Supportive Presence	--	-.53***	-.62***	.68***	.42***	-.46***	.47***	.56***	-.36***	.45***	.76***	.68***	-.55***	.82***	-.16***
2 Intrusiveness		--	.64***	-.40***	-.27***	.59***	-.32***	-.32***	.44***	-.39***	-.53***	-.50***	.46***	-.53***	.29***
3 Hostility			--	-.38***	-.24***	.75***	-.31***	-.35***	.49***	-.41***	-.65***	-.49***	.50***	-.59***	.15***
4 Quality of Instruction				--	.58***	-.38***	.55***	.46***	-.32***	.46***	.62***	.49***	-.50***	.65***	-.36***
5 Confidence					--	-.20***	.38***	.26***	-.29***	.30***	.38***	.27***	-.33***	.44***	-.43***
6 Harsh Treatment						--	-.33***	-.34***	.50***	-.43***	-.57***	-.43***	.47***	-.48***	.21***
Ratings of the Child															
7 Persistence							--	.65***	-.59***	.81***	.66***	.56***	-.66***	.64***	-.48***
8 Enthusiasm								--	-.43***	.56***	.73***	.76***	-.62***	.70***	-.19***
9 Negativity									--	-.75***	-.60***	-.50***	.66***	-.53***	.52***
10 Compliance										--	.68***	.54***	-.75***	.60***	-.54***
11 Experience of Session											--	.76***	-.74***	.85***	-.33***
12 Affection to Mother												--	-.68***	.80***	-.17***
13 Avoidance of Mother													--	-.70***	.38***
Ratings of the Relationship															
14 Quality of Relationship														--	-.29***
15 Boundary Dissolution															--
Sample size	290														

Source: New Chance coded observational study variables.

Notes: Statistical significance levels are indicated as *** <= 1 percent.

Calculations for this table used data for all 290 respondents for whom there were observational data coded for affect and for whom there were 18-month follow-up survey data, including those with values of zero for outcomes and New Chance enrollees (i.e., experimentals) who did not participate in the program. The actual sample size in this table is 284 because of missing data due to videotape problems.

BEST COPY AVAILABLE

Table 4.3

Reliability: Intraclass Correlations and Kappa

Observational Variable	Intraclass Correlation
Ratings of the Mother	
Supportive Presence	0.811
Intrusiveness	0.800
Hostility	0.810
Quality of Instruction	0.792
Ratings of the Child	
Persistence	0.835
Enthusiasm	0.838
Negativity	0.861
Compliance	0.830
Experience of Session	0.832
Affection to Mother	0.800
Avoidance of Mother	0.818
Ratings of the Relationship	
Quality of Relationship	0.823
Boundary Dissolution	0.757
Ratings of the Mother	Cohen's Kappa
Confidence	0.291

Table 4.4

Descriptive Statistics for Affective Quality Scales

Observational Variable	Range	Mean	SD
Ratings of the Mother			
Supportive Presence	1-7	4.10	1.31
Intrusiveness	1-7	2.93	1.37
Hostility	1-7	2.19	1.36
Quality of Instruction	1-7	3.83	1.16
Confidence	1-3	2.01	0.75
Harsh Treatment	0-7	0.38	0.96
Ratings of the Child			
Persistence	1-7	4.47	1.18
Enthusiasm	1-7	4.25	1.25
Negativity	1-7	2.08	1.36
Compliance	1-7	4.52	1.28
Experience of Session	1-7	4.27	1.26
Affection to Mother	1-7	4.31	1.24
Avoidance of Mother	1-7	2.23	1.53
Ratings of the Relationship			
Quality of Relationship	1-7	4.03	1.36
Boundary Dissolution	1-7	2.34	1.29
Sample size	290		

Source: New Chance coded observational study variables.

Notes: Calculations for this table used data for all 290 respondents for whom there were observational data coded for affect and for whom there were 18-month follow-up survey data, including those with values of zero for outcomes and New Chance enrollees (i.e., experimentals) who did not participate in the program. The actual sample size in this table is 284 because of missing data due to videotape problems.

B. Baseline Subgroup Scores on the Affective Quality Scales

Table 4.5 presents means on the five scales chosen for detailed analyses for subgroups formed from demographic and psychosocial variables measured at baseline, prior to random assignment to the New Chance experimental group or the control group. This table describes how different subgroups of the sample score on the affective quality scales. Since the affective quality scale scores were measured approximately 21 months after assignment to the New Chance Program, the means in Table 4.5 have been statistically adjusted to remove any effects related to being in New Chance. The remainder of this section highlights some of the significant findings for the subgroups. The analyses consider relations between baseline subgroups and the affective quality scales. Although we provide interpretations for most of the findings, it is important to keep two caveats in mind while thinking about the subgroup results in this chapter: (1) we provide only one or two of many possible explanations for the findings, and (2) the subgroup variables may be confounded with each other or with unmeasured variables that would hinder interpretation. To guard against interpreting findings that may be due to chance we will limit our discussion of results in this section to subgroups where there is a significant relation to more than one affective quality scale.

1. **Demographic Characteristics: Number of Children**. Surprisingly, mothers with only one child at the beginning of the study treated their children more harshly and had less compliant children than mothers who had more than one child. These results may be due to a difference in mothers' childrearing experience. Mothers who have more than one child have had more experience raising children and thus may be more effective at limit setting and discipline, and have children who are more open to direction. It is also possible that mothers with more than one child were observed with older focal children, and this finding may reflect a child age difference (see "child's age" in Table 4.6).

Race/Ethnicity. There are significant differences by race of mother on several of the scales. At this time we have insufficient evidence to decide whether these findings represent real differences between the subgroups or artifacts of our methodology. One possibility is that race of mother is confounded with some other variables that lead to more negative outcomes during the observation. Consequently, we have decided to control for race of mother in subsequent analyses.

2. **Education and Literacy: TABE Reading Score.** Mothers who had high reading levels at baseline are more supportive and less harsh with their children than mothers with low reading levels. There appears to be a relation between level of literacy and the affective quality scales, despite the fact that there was no relation between highest grade completed (coded as 10th or below/11th or above) and any of the scales. Higher reading levels at baseline may have resulted in the mothers being more comfortable interacting with their children in a teaching context. Mothers with lower reading levels may have been more stressed by the session.

3. **Psychosocial Characteristics: Satisfaction with Emotional Support.** Mothers who endorsed complete satisfaction with emotional support received from others, i.e., a five on a five-point scale, were actually less supportive toward their children and scored lower on Quality of Relationship than mothers who endorsed anything less than a five. While this finding may

Table 4.5

Means on Teaching Tasks Ratings, by Subgroups Formed at Baseline

Subgroup Formed at Baseline	n	Mother's Supportive Presence	Mother's Harsh Treatment	Child's Compliance	Child's Affection to Mother	Quality of Relationship
Full Sample						
Mean		4.10	0.38	4.52	4.31	4.03
Standard deviation		1.31	0.96	1.28	1.24	1.36
Observed range		1-7	0-6	1-7	1-7	1-7
Demographic Characteristics						
Gender of child[a]						
Female	141	4.08	0.36	4.60	4.27	4.03
Male	143	4.12	0.40	4.43	4.35	4.03
Number of children						
1	180	4.01	0.46 *	4.38 **	4.26	3.92
2 or more	104	4.26	0.24	4.76	4.39	4.22
Race/Ethnicity						
Black	238	3.95 ***	0.43 **	4.49	4.21 ***	3.92 ***
White	46	4.88	0.12	4.65	4.83	4.60
Age of mother (years)						
16-17	54	4.03	0.34	4.39	4.36	3.90
18-19	146	4.01	0.39	4.54	4.23	3.94
20-22	84	4.30	0.39	4.55	4.40	4.26
Living arrangement (mother)						
Living with mother	102	3.94	0.38	4.45	4.34	3.95
Not living with mother	175	4.20	0.38	4.56	4.29	4.05
Living arrangement (partner/husband)						
Living with partner/husband	28	4.30	0.16	4.33	4.45	4.09
Not living with partner/husband	249	4.08	0.40	4.54	4.29	4.01

(continued)

Table 4.5 (continued)

Subgroup Formed at Baseline	n	Mother's Supportive Presence	Mother's Harsh Treatment	Child's Compliance	Child's Affection to Mother	Quality of Relationship
Age at first child's birth (years)						
13-16	131	4.16	0.31	4.56	4.35	4.05
17-19	153	4.05	0.44	4.48	4.27	4.01
Age of youngest child (years)						
Under 1	106	3.95	0.32	4.38	4.10 **	3.83 *
1 or older	178	4.19	0.42	4.60	4.43	4.14
Number of pregnancies						
1 or 2	213	4.11	0.36	4.55	4.36	4.02
3 or more	70	4.09	0.42	4.45	4.19	4.07
Education and Literacy						
Highest grade completed						
10th or below	182	4.12	0.42	4.58	4.32	4.04
11th or above	102	4.07	0.32	4.40	4.28	4.00
Interval since last attended regular high school						
2 years or less	130	4.03	0.45	4.42	4.27	3.95
3 years or more	147	4.16	0.32	4.62	4.35	4.08
TABE reading test score (grade equivalent)^b						
7th grade or below	130	3.82 ***	0.47 *	4.55	4.17	3.84
8th or 9th grade	78	4.27	0.43	4.51	4.43	4.21
10th grade or above	76	4.42	0.18	4.48	4.42	4.16

(continued)

Table 4.5 (continued)

Subgroup Formed at Baseline	n	Mother's Supportive Presence	Mother's Harsh Treatment	Child's Compliance	Child's Affection to Mother	Quality of Relationship
Employment and AFDC History						
Ever employed						
Yes	221	4.14	0.38	4.53	4.30	4.05
No	63	3.96	0.39	4.47	4.33	3.95
Family received AFDC when sample member was growing up						
Never	82	4.19	0.29	4.74	4.52	4.13
Sometimes	142	4.17	0.40	4.45	4.29	4.04
Always	59	3.83	0.41	4.40	4.07	3.88
Psychosocial Characteristics						
CES-D (depression) score[c]						
0-15 (not at risk)	137	4.13	0.38	4.43	4.33	4.02
16-23 (at some risk)	76	4.16	0.37	4.78	4.34	4.19
24-60 (at high risk)	71	3.99	0.41	4.41	4.23	3.88
Self-Esteem score[d]						
Below mean	74	4.15	0.28	4.57	4.51	4.19
At or above mean (35)	210	4.08	0.42	4.50	4.24	3.97
Number of sources of emotional support						
0-2	127	4.11	0.39	4.48	4.29	4.01
3 or more	157	4.09	0.37	4.55	4.32	4.04
Level of satisfaction with emotional support[e]						
Less than very satisfied	135	4.33 ***	0.33	4.56	4.44	4.30 ***
Very satisfied	149	3.89	0.43	4.47	4.19	3.78

(continued)

Table 4.5 (continued)

Subgroup Formed at Baseline	n	Mother's Supportive Presence	Mother's Harsh Treatment	Child's Compliance	Child's Affection to Mother	Quality of Relationship
Locus of Control score[f]						
Below mean	111	3.89 **	0.55 **	4.41	4.16 *	3.79 **
At or above mean (22)	173	4.24	0.28	4.58	4.41	4.18
Sample size	290					

Sources: New Chance baseline enrollment data and coded observational study variables.

Notes: Means are adjusted for experimental/control status in order to be descriptive of the overall sample.

For subgroups that have more than two levels, the placement of asterisks does not indicate where the significant difference lies, but only that there was a significant difference among the set of means.

Statistical significance levels are indicated as *** <= 1 percent, ** <= 5 percent, * <= 10 percent, which refers to the probability that sample means are different from each other only because of random error.

Mother's Supportive Presence, Child's Compliance, Child's Affection to Mother, and Quality of Relationship scales can range from 1 (low) to 7 (high). Mother's Harsh Treatment is a summary score that combines 6 infrequently seen checklist behaviors and a high score (4-7) on the Mother's Hostility scale. It can range from 0 (no items) to 7 (all items).

Calculations for this table used data for all 290 respondents for whom there were observational data coded for affect and for whom there were 18-month follow-up survey data, including those with values of zero for outcomes and New Chance enrollees (i.e., experimentals) who did not participate in the program. The actual sample size in this table is 284 because of missing or unusable items from some respondents' questionnaires and videotape problems.

[a] When a sample member had more than one child, her response refers to the randomly selected focal child.

[b] The test used to measure reading ability was the reading part of the Tests of Adult Basic Education (TABE). Most sites administered the survey form of the test, but some administered the full reading test.

[c] The Center for Epidemiological Studies Depression (CES-D) Scale is a widely used measure of depression; scores can range from 0 to 60.

[d] The measure of self-esteem used was the Rosenberg Self-Esteem Scale, a 10-item scale that assesses a person's global sense of self-worth. Scores can range from 10 to 50; 30 is considered the neutral midpoint.

[e] Enrollees were asked about their degree of satisfaction with the emotional support ("people who listen to you, reassure you, and show you they care") they received. Levels range from 1 (very dissatisfied) to 5 (very satisfied).

[f] The Locus of Control Scale is a 6-item adaptation of the longer scale originally developed by Julien Rotter (1966). Scores can range from 6 to 30; 18 is considered the neutral midpoint.

131

135

Table 4.6

Means on Teaching Tasks Ratings, by 18-Month Follow-Up Characteristics

Subgroup Formed at 18 Months	n	Mother's Supportive Presence	Mother's Harsh Treatment	Child's Compliance	Child's Affection to Mother	Quality of Relationship
Full Sample						
Mean		4.10	0.38	4.52	4.31	4.03
Standard deviation		1.31	0.96	1.28	1.24	1.36
Observed range		1-7	0-6	1-7	1-7	1-7
Demographic Characteristics						
Child's age at observational study (months)						
24-36	83	3.60 ***	0.70 ***	3.83 ***	3.73 ***	3.37 ***
37-47	114	4.19	0.35	4.54	4.37	4.10
48-63	87	4.46	0.12	5.14	4.78	4.56
Residence						
Living with partner/husband	53	4.30	0.37	4.45	4.45	4.17
Not living with partner/husband	231	4.06	0.39	4.53	4.28	3.99
Living with parent/grandparent	99	3.99	0.37	4.54	4.26	3.98
Not living with parent/grandparent	185	4.16	0.39	4.51	4.33	4.05
Education and Literacy						
TABE reading score (grade equivalent)						
7th grade or below	135	3.73 ***	0.53 **	4.51	4.13 **	3.75 ***
8th or 9th grade	70	4.39	0.30	4.50	4.38	4.21
10th grade or above	38	4.16	0.20	4.23	4.58	4.18
Mother's education at 18 months						
No high school diploma or GED, not in school	148	3.94	0.50	4.44	4.23	3.89
In process of getting GED or high school diploma	29	4.01	0.19	4.50	4.14	3.94
Has GED	54	4.35	0.27	4.61	4.44	4.24
High school/trade school diploma or attending college	53	4.35	0.27	4.65	4.49	4.25

(continued)

Table 4.6 (continued)

Subgroup Formed at 18 Months	n	Mother's Supportive Presence	Mother's Harsh Treatment	Child's Compliance	Child's Affection to Mother	Quality of Relationship
Earnings from Employment						
Respondent's total earnings over the 18-month period[a]						
No earnings	164	3.98 *	0.43	4.42	4.30	3.91
Below mean	84	4.14	0.32	4.60	4.29	4.10
At or above mean ($3,323.87)	36	4.54	0.31	4.78	4.41	4.41
Total earnings from respondent and partner or others during month 18[a]						
No earnings	208	3.92 ***	0.44	4.39 **	4.22	3.87 ***
Below mean	42	4.42	0.27	4.85	4.50	4.40
At or above mean ($719.40)	34	4.79	0.17	4.88	4.61	4.52
Fertility						
Pregnancy during follow-up	161	3.95 **	0.37	4.46	4.21	3.90 *
No pregnancy during follow-up	123	4.30	0.40	4.59	4.44	4.20
Birth during follow-up	100	3.91 *	0.42	4.42	4.25	3.91
No birth during follow-up	184	4.20	0.36	4.57	4.34	4.09
Abortion during follow-up	41	3.90	0.29	4.64	4.17	3.80
No abortion during follow-up	243	4.14	0.40	4.50	4.33	4.07
Two or more pregnancies during follow-up	40	4.07	0.22	4.32	4.34	3.94
Did not have two or more pregnancies during follow-up	244	4.11	0.41	4.55	4.30	4.04
Sexually active at follow-up	238	4.11	0.33	4.55	4.31	4.04
Not sexually active at follow-up	44	4.13	0.52	4.35	4.35	3.97
Child Care[b]						
Any child care during follow-up	248	4.12	0.39	4.56 *	4.31	4.04
No child care during follow-up	26	3.93	0.45	4.09	4.32	3.79

(continued)

Table 4.6 (continued)

Subgroup Formed at 18 Months	n	Mother's Supportive Presence	Mother's Harsh Treatment	Child's Compliance	Child's Affection to Mother	Quality of Relationship
Child ever in child care before age 1	82	3.99	0.41	4.49	4.12	3.82
Child not in care before age 1	192	4.15	0.39	4.52	4.39	4.11
Number of months child was in day care/preschool						
Below mean	179	4.04	0.42	4.50	4.22	3.89 **
At or above mean (3.7 months)	105	4.21	0.33	4.54	4.46	4.26
Number of months child was in care by grandparent						
Below mean	198	4.08	0.41	4.42 *	4.33	3.97
At or above mean (2.9 months)	86	4.15	0.32	4.73	4.26	4.17
Psychosocial characteristics						
CES-D (depression) score[c]						
0-15 (not at risk)	173	4.16	0.30	4.58 *	4.37	4.10
16-23 (at some risk)	65	4.13	0.48	4.65	4.32	4.03
24-60 (at high risk)	46	3.85	0.55	4.10	4.07	3.74
Mastery score[d]						
Below mean	115	4.01	0.46	4.43	4.22	3.98
At or above mean (22)	169	4.16	0.33	4.57	4.37	4.06
Difficult Life Circumstances score[e]						
Below mean	140	4.27 **	0.36	4.52	4.45 *	4.13
At or above mean (2.8)	144	3.94	0.41	4.52	4.17	3.93
Number of social supports[f]						
0-2	200	4.04	0.44	4.39 **	4.23 *	3.94 *
3 or more	83	4.26	0.25	4.81	4.50	4.23
Contact with Father of Child						
Father babysat for child in past year	125	3.99	0.34	4.53	4.22	3.97
Father did not babysit for child in past year	158	4.19	0.40	4.52	4.39	4.09

(continued)

Table 4.6 (continued)

Subgroup Formed at 18 Months	n	Mother's Supportive Presence	Mother's Harsh Treatment	Child's Compliance	Child's Affection to Mother	Quality of Relationship
Father took child overnight	137	3.96 *	0.40	4.55	4.18 *	3.93
Father did not take child overnight	146	4.24	0.35	4.51	4.44	4.13
Father saw child once or not at all	80	4.15	0.47	4.51	4.20	3.94
Father saw child more than once	203	4.09	0.34	4.53	4.36	4.07
Substance Abuse						
Respondent ever high on alcohol in past month	104	4.00	0.27	4.55	4.21	3.94
Respondent never high on alcohol in past month	176	4.16	0.46	4.49	4.36	4.07
Respondent used drugs in past month[g]	43	3.68 **	0.38	4.43	4.24	3.84
Respondent did not use drugs in past month[g]	236	4.19	0.39	4.53	4.32	4.07
Sample size	290					

Sources: New Chance 18-month follow-up survey and coded observational study variables.

Notes: Means are adjusted for experimental/control status in order to be descriptive of the overall sample.

A two-tailed F-test was applied to each difference. Statistical significance levels are indicated as *** <= 1 percent, ** <= 5 percent, * <= 10 percent, which refers to the probability that sample means are different from each other only because of random error.

For subgroups that have more than two levels, the placement of asterisks does not indicate where the significant difference lies, but only that there was a significant difference among the set of means.

Mother's Supportive Presence, Child's Compliance, Child's Affection to Mother, and Quality of Relationship scales can range from 1 (low) to 7 (high). Mother's Harsh Treatment is a summary score that combines 6 infrequently seen checklist behaviors and a high score (4-7) on the Mother's Hostility scale. It can range from 0 (no items) to 7 (all items).

Calculations for this table used data for all 290 respondents for whom there were observational data coded for affect and for whom there were 18-month follow-up survey data, including those with values of zero for outcomes and New Chance enrollees (i.e., experimentals) who did not participate in the program. The actual sample size in this table is 284 because of missing or unusable items from some respondents' questionnaires and videotape problems.

[a] Means were computed excluding values of zero.

[b] Regular child care was defined as an ongoing arrangement used while the mother was in school, in training, or working.

[c] The Center for Epidemiological Studies Depression (CES-D) Scale is a widely used measure of depression; scores can range from 0 to 60.

[d] The Mastery Scale measures sense of mastery over personal events. Scores can range from 7 to 28.

[e] Scores represent total numbers of ongoing problems or stresses the respondent faces, of a list of 10 problems.

[f] Respondents were asked to indicate the total number of individuals to whom they could turn for support.

[g] Self-report measures that included use of marijuana, cocaine, crack, heroin, PCP, and ice.

seem counterintuitive, we believe that it may reflect unrealistic assessments by some of the mothers of their difficult situations. This same lack of realism may later interfere with their ability to adapt and interact successfully with their children in a somewhat stressful situation.

Locus of Control. Mothers who felt that they had more control over their own lives were more supportive, less harsh, had children who showed more affection, and were rated higher on the Quality of Relationship scale than mothers who felt that they had little control over their own lives. Mothers who feel more in control of their lives at baseline may feel more confident and in control of the Teaching Task session as well. This confidence may allow them to navigate the challenges of the session and maintain positive interactions with their children without resorting to coercive strategies.

4. Summary: The findings from Table 4.5 show that some maternal qualities at baseline were associated with more successful negotiation of the Teaching Task sessions about 21 months later. Mothers who had a realistic yet confident outlook at baseline may have been better able to navigate the sessions without allowing the stress of the session to interfere with their relationships. Interestingly enough, there was *no* relation between the affective quality scales and either the living arrangement variables or most of the psychosocial variables (notably depression).

C. 18-month Subgroup Scores on the Affective Quality Scales

Table 4.6 presents means, on the five scales chosen for detailed analyses, for subgroups formed from demographic and psychosocial variables measured at the 18-month follow-up visit. The follow-up visit occurred about three months before the Teaching Task observations (which took place on average at 21 months) and are therefore more contemporaneous with the observations than the baseline measures. Like Table 4.5, this table describes how different subgroups of the sample score on the affective quality scales. The means in Table 4.6 have also been statistically adjusted to remove any effects related to being in the New Chance Program from the scales. It is important to note, however, that this statistical adjustment cannot correct for any effects that being assigned to the New Chance Program experimental or control groups might have had in determining mothers' membership in particular 18-month subgroups. The remainder of this section highlights some of the significant findings for the subgroups. Findings for the 18-month subgroup analyses are subject to the same two caveats put forth at the beginning of the previous section: (1) we provide only one or two of many possible explanations, and (2) subgroup variables may be confounded with each other or with unmeasured variables. To guard against interpreting findings that may be due to chance we again limit our discussion of results in this section to subgroups where there is a significant relation to more than one affective quality scale.

1. Demographic Characteristics: Child's Age at Observation. There was a strong and clear relation between the age of the children in the observational session and their ratings on all of the scales: if the children were older, the affective quality of the relationship was better. This finding is not surprising given that the Teaching Task session was designed to assess relationships between mothers and their 42-month-old children. Younger children have more difficulty with the tasks and increase the demands placed on their mothers, as compared to older

141

children, who find the session easier. These age-related differences certainly affect the quality of mother-child interactions during the task session. Therefore, we have controlled for child's age in subsequent impact analyses.

2. **Education and Literacy: TABE Reading Score.** Mothers were less supportive and more harsh and children were less affectionate in dyads where the mother's grade equivalent TABE reading score was 7th grade or below. These dyads also had lower Quality of Relationship ratings than dyads where the mother had a TABE score above 7th grade. While there were small differences between the two groups with higher literacy scores, both groups had more positive mother-child interactions than the lowest group. Low literacy seems to be a risk factor for lower affective quality of mother-child interactions.

3. **Earnings from Employment: Total Earnings for Mother and Partner or Others During Month 18.** In dyads where the total earnings for the mother and partner or others during month 18 (the month the interview took place) were higher, the mothers were more supportive, their children were more compliant, and their Quality of Relationship ratings were higher than in dyads where total earnings were lower. This finding may be an indication that when the mothers' lives are going better, that is, they and/or a partner are employed in a better paying job, we see fewer disruptions in mother-child interactions.

4. **Psychosocial Characteristics: Difficult Life Circumstances Score.** Mothers who reported having fewer difficult life circumstances were more supportive and their children were more affectionate. This finding may be evidence supporting McLoyd's suggestion (1990) that reduced stress in the mothers' lives can have a beneficial effect on the affective quality of mother-child interactions.

Number of Social Supports. Mothers who reported having at least three sources of social support had higher Quality of Relationship ratings and their children were more compliant and affectionate. As has been suggested in previous research, mothers who experience more support may have less stress and consequently have more positive interactions with their children (Crnic et al., 1983).

5. **Summary**. Mothers who reported less stress and better economic and interpersonal circumstances at the 18-month follow-up had more positive interactions with their children during the Teaching Task sessions approximately three months later.

D. New Chance Impacts on the Affective Quality Scales

Table 4.7 shows whether participation in New Chance had impacts on the affective quality of mother-child interactions in the Teaching Task session. There was a difference between the New Chance experimental and control groups on one of the scales: Mother's Harsh Treatment. Mothers in the New Chance experimental group displayed significantly fewer harsh behaviors toward their children in the Teaching Task session than mothers in the control group.

E. New Chance Impacts for Baseline Subgroups

The findings in Tables 4.8 through 4.12 allow us to determine whether participation in the New Chance Program had impacts on the affective quality scales for some subgroups more

Table 4.7

Impacts of New Chance on Teaching Tasks Ratings

Observational Variable	Experimentals	Controls	Difference	p[a]
Ratings of the Mother				
Supportive Presence	4.16	4.00	0.16	0.300
Harsh Treatment	0.29	0.55	-0.26 **	0.026
Ratings of the Child				
Compliance	4.48	4.58	-0.10	0.486
Affection to Mother	4.35	4.24	0.11	0.483
Ratings of the Relationship				
Quality of Relationship	4.10	3.90	0.20	0.208
Sample size	184	106		

Source: New Chance coded observational study variables.

Notes: Calculations for this table used data for all 290 respondents for whom there were observational data coded for affect and 18-month follow-up survey data, including those with values of zero for outcomes and New Chance enrollees (i.e., experimentals) who did not participate in the program. The actual sample size in this table is 284 (183 experimentals, 101 controls) because of missing data due to videotape problems.

The averages are adjusted using linear analysis of covariance procedures controlling for seven kinds of difference in characteristics before random assignment (age of child, maternal literacy, whether mother had more than one child, gender of child, race/ethnicity, Philadelphia site, Portland site).

[a] An F-test was applied to each difference in regression-adjusted means. The column labeled "p" is the statistical significance level of each between-group impact. That is, p is the probability that the averages for the experimental and control groups are different from each other only because of random error. Statistical significance levels are indicated as ***<=1 percent; **<=5 percent; *<= 10 percent.

Table 4.8

Impacts of New Chance on Mother's Supportive Presence, by Subgroups Formed at Baseline

Characteristics and Subgroup at Baseline	n	Mother's Supportive Presence		Within-Subgroup Impact	p[a]	Between-Subgroups Difference	p[b]
		Experimentals	Controls				
Full sample							
Mean		4.16	4.00	0.16	0.300		
Demographic Characteristics							
Child's age at observational study (months)						--	0.531
24-36	83	3.75	3.51	0.24	0.393		
37-47	114	4.27	4.01	0.26	0.267		
48-63	87	4.37	4.48	-0.11	0.677		
Gender of child[c]						0.40	0.182
Female	141	4.17	3.82	0.35 *	0.095		
Male	143	4.15	4.20	-0.05	0.808		
Number of children						0.25	0.425
1	180	4.16	3.90	0.26	0.191		
2 or more	104	4.15	4.14	0.01	0.976		
Race/ethnicity						0.32	0.500
Black	238	4.05	3.86	0.19	0.230		
White	46	4.72	4.85	-0.13	0.774		
Age of mother (years)						--	0.883
16-17	54	4.41	4.12	0.29	0.405		
18-19	146	4.05	3.90	0.15	0.498		
20-22	84	4.17	4.10	0.07	0.807		
Living arrangement (mother)						-0.32	0.321
Living with mother	102	4.08	4.11	-0.03	0.906		
Not living with mother	175	4.21	3.92	0.29	0.152		
Living arrangement (partner/husband)						-0.21	0.697
Living with partner/husband	28	4.04	4.07	-0.03	0.953		
Not living with partner/husband	249	4.18	4.00	0.18	0.265		
Age at first child's birth (years)						0.00	0.986
13-16	131	4.24	4.08	0.16	0.456		
17-19	153	4.09	3.93	0.16	0.456		
Age of youngest child (years)						0.26	0.399
Under 1	106	4.11	3.80	0.31	0.199		
1 or older	178	4.19	4.14	0.05	0.817		
Number of pregnancies						0.55	0.116
1 or 2	213	4.25	3.97	0.28	0.114		
3 or more	70	3.87	4.14	-0.27	0.373		

(continued)

Table 4.8 (continued)

Characteristics and Subgroup at Baseline	n	Mother's Supportive Presence		Within-Subgroup Impact	p[a]	Between-Subgroups Difference	p[b]
		Experimentals	Controls				
Education and Literacy							
Highest grade completed						-0.05	0.854
10th or below	182	4.21	4.08	0.13	0.507		
11th or above	102	4.05	3.87	0.18	0.452		
Interval since last attended regular high school						0.26	0.391
2 years or less	130	4.25	3.96	0.29	0.192		
3 years or more	147	4.06	4.03	0.03	0.882		
TABE reading test score (grade equivalent)[d]						--	0.277
7th grade or below	130	3.98	4.08	-0.10	0.675		
8th or 9th grade	78	4.30	3.92	0.38	0.181		
10th grade or above	76	4.35	3.95	0.40	0.163		
Employment and AFDC History							
Ever employed						0.11	0.744
Yes	221	4.23	4.05	0.18	0.291		
No	63	3.90	3.83	0.07	0.848		
Family received AFDC when sample member was growing up						--	0.904
Never	82	4.16	3.91	0.25	0.402		
Sometimes	142	4.22	4.10	0.12	0.564		
Always	59	3.97	3.91	0.06	0.843		
Psychosocial Characteristics							
CES-D (depression) score[e]						--	0.576
0-15 (not at risk)	137	4.11	3.95	0.16	0.452		
16-23 (at some risk)	76	4.18	4.23	-0.05	0.847		
24-60 (at high risk)	71	4.21	3.83	0.38	0.209		
Self-Esteem score[f]						0.13	0.693
Below mean	74	4.27	4.01	0.26	0.386		
At or above mean (35)	210	4.12	3.99	0.13	0.480		
Number of sources of emotional support						-0.13	0.674
0-2	127	4.15	4.06	0.09	0.674		
3 or more	157	4.16	3.94	0.22	0.293		

(continued)

Table 4.8 (continued)

Characteristics and Subgroup at Baseline	n	Mother's Supportive Presence Experimentals	Controls	Within-Subgroup Impact	p[a]	Between-Subgroups Difference	p[b]
Level of satisfaction with emotional support[g]						0.14	0.622
Less than very satisfied	135	4.44	4.19	0.25	0.235		
Very satisfied	149	3.92	3.81	0.11	0.589		
Locus of Control score[h]						0.02	0.923
Below mean	111	4.09	3.92	0.17	0.461		
At or above mean (22)	173	4.20	4.05	0.15	0.451		
Sample size		184	106				

Sources: New Chance baseline enrollment data and coded observational study variables.

Notes: Calculations for this table used data for all 290 respondents for whom there were observational data coded for affect and for whom there were 18-month follow-up survey data, including those with values of zero for outcomes and New Chance enrollees (i.e., experimentals) who did not participate in the program. The actual sample sizes in this table are 183 experimentals and 101 controls because of missing or unusable items from some respondents' questionnaires and videotape problems.

The averages are adjusted using linear analysis of covariance procedures controlling for up to seven kinds of difference in characteristics before random assignment (age of child, maternal literacy, whether mother had more than one child, gender of child, race/ethnicity, Philadelphia site, and Portland site).

[a] A two-tailed t-test was applied to each difference in regression-adjusted means. "P" is the statistical significance level of each within-subgroup impact. That is, p is the probability that sample estimates are different from zero or from each other only because of random error. Statistical significance levels are indicated as ***<=1 percent; **<=5 percent; *<= 10 percent.

[b] An F-test was applied to the interaction between the subgroup and experimental/control status. "P" is the statistical significance level of each between-subgroups difference in impacts. That is, p is the probability that sample estimates are different from zero or from each other only because of random error. Statistical significance levels are indicated as ***<=1 percent; **<=5 percent; *<= 10 percent.

[c] When a sample member had more than one child, her response refers to the randomly selected focal child.

[d] The test used to measure reading ability was the reading part of the Tests of Adult Basic Education (TABE). Most sites administered the survey form of the test, but some administered the full reading test.

[e] The Center for Epidemiological Studies Depression (CES-D) Scale is a widely used measure of depression; scores can range from 0 to 60.

[f] The measure of self-esteem used was the Rosenberg Self-Esteem Scale, a 10-item scale that assesses a person's global sense of self-worth. Scores can range from 10 to 50; 30 is considered the neutral midpoint.

[g] Enrollees were asked about their degree of satisfaction with the emotional support ("people who listen to you, reassure you, and show you they care") they received. Levels range from 1 (very dissatisfied) to 5 (very satisfied).

[h] The Locus of Control Scale is a 6-item adaptation of the longer scale originally developed by Julien Rotter (1966). Scores can range from 6 to 30; 18 is considered the neutral midpoint.

Table 4.9

Impacts of New Chance on Mother's Harsh Treatment, by Subgroups Formed at Baseline

Characteristics and Subgroup at Baseline	n	Mother's Harsh Treatment Experimentals	Mother's Harsh Treatment Controls	Within-Subgroup Impact	p[a]	Between-Subgroups Difference[b]	p[c]
Full Sample							
Mean		0.29	0.55	-0.26 **	0.026		
Demographic Characteristics							
Child's age at observational study (months)						-- *	0.072
24-36	83	0.43	1.04	-0.61 ***	0.004		
37-47	114	0.29	0.51	-0.22	0.210		
48-63	87	0.19	0.13	0.05	0.778		
Gender of child[d]						-0.31	0.177
Female	141	0.23	0.64	-0.40 **	0.011		
Male	143	0.35	0.45	-0.09	0.560		
Number of children						-0.01	0.996
1	180	0.34	0.60	-0.26 *	0.083		
2 or more	104	0.21	0.47	-0.25	0.162		
Race/ethnicity						-0.37	0.302
Black	238	0.32	0.62	-0.30 **	0.014		
White	46	0.14	0.06	0.07	0.827		
Age of mother (years)						--	0.570
16-17	54	0.11	0.23	-0.12	0.631		
18-19	146	0.33	0.51	-0.18	0.255		
20-22	84	0.36	0.80	-0.43 **	0.040		
Living arrangement (mother)						0.15	0.523
Living with mother	102	0.23	0.39	-0.16	0.389		
Not living with mother	175	0.33	0.64	-0.31 **	0.037		
Living arrangement (partner/husband)						0.58	0.150
Living with partner/husband	28	0.31	0.01	0.29	0.452		
Not living with partner/husband	249	0.29	0.59	-0.29 **	0.016		
Age at first child's birth (years)						-0.01	0.964
13-16	131	0.21	0.48	-0.26	0.109		
17-19	153	0.36	0.62	-0.25	0.111		
Age of youngest child (years)						-0.20	0.408
Under 1	106	0.10	0.49	-0.38 **	0.036		
1 or older	178	0.40	0.59	-0.18	0.212		
Number of pregnancies						0.27	0.308
1 or 2	213	0.27	0.44	-0.16	0.206		
3 or more	70	0.36	0.80	-0.43 *	0.057		

(continued)

BEST COPY AVAILABLE

Table 4.9 (continued)

Characteristics and Subgroup at Baseline	n	Mother's Harsh Treatment Experimentals	Mother's Harsh Treatment Controls	Within-Subgroup Impact	p[a]	Between-Subgroups Difference[b]	p[c]
Education and Literacy							
Highest grade completed						-0.03	0.893
10th or below	182	0.29	0.56	-0.27 *	0.065		
11th or above	102	0.30	0.54	-0.24	0.202		
Interval since last attended regular high school						-0.13	0.602
2 years or less	130	0.28	0.61	-0.33 *	0.055		
3 years or more	147	0.31	0.52	-0.20	0.196		
TABE reading test score (grade equivalent)[e]						-- **	0.015
7th grade or below	130	0.43	0.62	-0.19	0.271		
8th or 9th grade	78	0.17	0.96	-0.79 ***	0.000		
10th grade or above	76	0.15	0.10	0.04	0.829		
Employment and AFDC History							
Ever employed						-0.29	0.295
Yes	221	0.25	0.58	-0.32 **	0.014		
No	63	0.44	0.47	-0.03	0.901		
Family received AFDC when sample member was growing up						--	0.234
Never	82	0.18	0.70	-0.52 **	0.021		
Sometimes	142	0.32	0.51	-0.19	0.228		
Always	59	0.41	0.38	0.03	0.896		
Psychosocial Characteristics							
CES-D (depression) score[f]						--	0.850
0-15 (not at risk)	137	0.30	0.59	-0.29 *	0.083		
16-23 (at some risk)	76	0.25	0.56	-0.31	0.164		
24-60 (at high risk)	71	0.32	0.47	-0.14	0.526		
Self-Esteem score[g]						0.18	0.496
Below mean	74	0.23	0.36	-0.12	0.581		
At or above mean (35)	210	0.32	0.62	-0.30 **	0.023		
Number of sources of emotional support						-0.05	0.839
0-2	127	0.28	0.56	-0.28 *	0.091		
3 or more	157	0.31	0.55	-0.23	0.134		

(continued)

Table 4.9 (continued)

Characteristics and Subgroup at Baseline	n	Mother's Harsh Treatment		Within-Subgroup Impact	p[a]	Between-Subgroups Difference[b]	p[c]
		Experimentals	Controls				
Level of satisfaction with emotional support[h]						0.23	0.317
Less than very satisfied	135	0.26	0.41	-0.14	0.378		
Very satisfied	149	0.33	0.70	-0.37 **	0.019		
Locus of Control score[i]						-0.02	0.936
Below mean	111	0.39	0.67	-0.27	0.135		
At or above mean (22)	173	0.23	0.48	-0.25 *	0.088		
Sample size		184	106				

Sources: New Chance baseline enrollment data and coded observational study variables.

Notes: Calculations for this table used data for all 290 respondents for whom there were observational data coded for affect and for whom there were 18-month follow-up survey data, including those with values of zero for outcomes and New Chance enrollees (i.e., experimentals) who did not participate in the program. The actual sample sizes in this table are 183 experimentals and 101 controls because of missing or unusable items from some respondents' questionnaires and videotape problems.

The averages are adjusted using linear analysis of covariance procedures controlling for up to seven kinds of difference in characteristics before random assignment (age of child, maternal literacy, whether mother had more than one child, gender of child, race/ethnicity, Philadelphia site, and Portland site).

[a] A two-tailed t-test was applied to each difference in regression-adjusted means. The column labeled "p" is the statistical significance level of each within-subgroup impact. That is, p is the probability that sample estimates are different from zero or from each other only because of random error. Statistical significance levels are indicated as ***<=1 percent; **<=5 percent; *<= 10 percent.

[b] For each characteristic with only two subgroups, the between-subgroups difference is the impact for the first subgroup less the impact for the second subgroup. For characteristics with more than two subgroups, no single difference between subgroup impacts can be calculated, as indicated by dashes in the table. However, it is possible to assess the statistical significance of variation across multiple subgroups, as indicated by the asterisks.

[c] An F-test was applied to the interaction between the subgroup and experimental/control status. The column labeled "p" is the statistical significance level of each between-subgroups difference in impacts. That is, p is the probability that sample estimates are different from zero or from each other only because of random error. Statistical significance levels are indicated as ***<=1 percent; **<=5 percent; *<= 10 percent.

[d] When a sample member had more than one child, her response refers to the randomly selected focal child.

[e] The test used to measure reading ability was the reading part of the Tests of Adult Basic Education (TABE). Most sites administered the survey form of the test, but some administered the full reading test.

[f] The Center for Epidemiological Studies Depression (CES-D) Scale is a widely used measure of depression; scores can range from 0 to 60.

[g] The measure of self-esteem used was the Rosenberg Self-Esteem Scale, a 10-item scale that assesses a person's global sense of self-worth. Scores can range from 10 to 50; 30 is considered the neutral midpoint.

[h] Enrollees were asked about their degree of satisfaction with the emotional support ("people who listen to you, reassure you, and show you they care") they received. Levels range from 1 (very dissatisfied) to 5 (very satisfied).

[i] The Locus of Control Scale is a 6-item adaptation of the longer scale originally developed by Julien Rotter (1966). Scores can range from 6 to 30; 18 is considered the neutral midpoint.

Table 4.10

Impacts of New Chance on Child's Compliance, by Subgroups Formed at Baseline

Characteristics and Subgroup at Baseline	n	Child's Compliance Experimentals	Child's Compliance Controls	Within-Subgroup Impact	p[a]	Between-Subgroups Difference	p[b]
Full Sample							
Mean		4.48	4.58	-0.10	0.486		
Demographic Characteristics							
Child's age at observational study (months)						--	0.187
24-36	83	3.84	3.90	-0.06	0.840		
37-47	114	4.57	4.44	0.13	0.594		
48-63	87	4.94	5.46	-0.52 *	0.053		
Gender of child[c]						0.14	0.626
Female	141	4.59	4.63	-0.04	0.864		
Male	143	4.37	4.55	-0.18	0.405		
Number of children						0.85 ***	0.005
1	180	4.51	4.28	0.23	0.216		
2 or more	104	4.39	5.01	-0.62 ***	0.009		
Race/ethnicity						0.37	0.418
Black	238	4.46	4.53	-0.07	0.697		
White	46	4.57	5.01	-0.44	0.319		
Age of mother (years)						--	0.465
16-17	54	4.62	4.48	0.14	0.674		
18-19	146	4.55	4.59	-0.04	0.836		
20-22	84	4.26	4.63	-0.37	0.173		
Living arrangement (mother)						0.10	0.720
Living with mother	102	4.55	4.57	-0.02	0.956		
Not living with mother	175	4.46	4.58	-0.12	0.522		
Living arrangement (partner/husband)						-1.41 ***	0.008
Living with partner/husband	28	3.82	5.19	-1.37 ***	0.007		
Not living with partner/husband	249	4.57	4.53	0.04	0.809		
Age at first child's birth (years)						-0.01	0.968
13-16	131	4.40	4.51	-0.11	0.600		
17-19	153	4.55	4.65	-0.10	0.627		
Age of youngest child (years)						-0.24	0.437
Under 1	106	4.32	4.57	-0.25	0.289		
1 or older	178	4.57	4.58	-0.01	0.947		
Number of pregnancies						0.19	0.575
1 or 2	213	4.57	4.65	-0.08	0.654		
3 or more	70	4.20	4.47	-0.27	0.366		

(continued)

Table 4.10 (continued)

Characteristics and Subgroup at Baseline	n	Child's Compliance		Within-Subgroup Impact	p[a]	Between-Subgroups Difference	p[b]
		Experimentals	Controls				
Education and Literacy							
Highest grade completed						0.10	0.747
10th or below	182	4.58	4.66	-0.08	0.675		
11th or above	102	4.29	4.47	-0.18	0.462		
Interval since last attended regular high school						0.42	0.162
2 years or less	130	4.55	4.44	0.11	0.607		
3 years or more	147	4.42	4.73	-0.31	0.137		
TABE reading test score (grade equivalent)[d]						--	0.974
7th grade or below	130	4.59	4.75	-0.16	0.475		
8th or 9th grade	78	4.39	4.47	-0.08	0.772		
10th grade or above	76	4.36	4.48	-0.12	0.684		
Employment and AFDC History							
Ever employed						-0.04	0.903
Yes	221	4.51	4.62	-0.11	0.496		
No	63	4.37	4.44	-0.07	0.823		
Family received AFDC when sample member was growing up						--	0.906
Never	82	4.70	4.73	-0.03	0.904		
Sometimes	142	4.42	4.58	-0.16	0.420		
Always	59	4.28	4.49	-0.21	0.498		
Psychosocial Characteristics							
CES-D (depression) score[e]						--	0.265
0-15 (not at risk)	137	4.48	4.38	0.10	0.657		
16-23 (at some risk)	76	4.57	5.05	-0.48 *	0.092		
24-60 (at high risk)	71	4.38	4.45	-0.07	0.816		
Self-Esteem score[f]						0.06	0.885
Below mean	74	4.54	4.60	-0.06	0.819		
At or above mean (35)	210	4.46	4.58	-0.12	0.499		
Number of sources of emotional support						-0.19	0.536
0-2	127	4.40	4.61	-0.21	0.344		
3 or more	157	4.54	4.56	-0.02	0.914		

(continued)

Table 4.10 (continued)

Characteristics and Subgroup at Baseline	n	Child's Compliance: Experimentals	Child's Compliance: Controls	Within-Subgroup Impact	p[a]	Between-Subgroups Difference	p[b]
Level of satisfaction with emotional support[g]						0.14	0.625
Less than very satisfied	135	4.59	4.62	-0.03	0.922		
Very satisfied	149	4.38	4.55	-0.17	0.420		
Locus of Control score[h]						-0.14	0.650
Below mean	111	4.46	4.64	-0.18	0.429		
At or above mean (22)	173	4.50	4.54	-0.04	0.795		
Sample size		184	106				

Sources and Notes: See Table 4.8.

Table 4.11

Impacts of New Chance on Child's Affection to Mother, by Subgroups Formed at Baseline

Characteristics and Subgroup at Baseline	n	Child's Affection to Mother: Experimentals	Child's Affection to Mother: Controls	Within-Subgroup Impact	p[a]	Between-Subgroup Difference	p[b]
Full Sample							
Mean		4.35	4.24	0.11	0.483		
Demographic Characteristics							
Child's age at observational study (months)						--	0.496
24-36	83	3.80	3.71	0.09	0.733		
37-47	114	4.45	4.18	0.27	0.258		
48-63	87	4.70	4.86	-0.16	0.556		
Gender of child[c]						0.36	0.229
Female	141	4.36	4.08	0.28	0.182		
Male	143	4.34	4.42	-0.08	0.702		
Number of children						0.12	0.679
1	180	4.37	4.22	0.15	0.421		
2 or more	104	4.30	4.27	0.03	0.907		
Race/ethnicity						0.49	0.287
Black	238	4.26	4.11	0.15	0.311		
White	46	4.79	5.13	-0.34	0.443		
Age of mother (years)						--	0.677
16-17	54	4.72	4.35	0.37	0.289		
18-19	146	4.23	4.21	0.02	0.922		
20-22	84	4.29	4.25	0.04	0.879		
Living arrangement (mother)						-0.24	0.437
Living with mother	102	4.46	4.48	-0.02	0.936		
Not living with mother	175	4.29	4.07	0.22	0.255		
Living arrangement (partner/husband)						-0.66	0.212
Living with partner/husband	28	4.11	4.60	-0.49	0.334		
Not living with partner/husband	249	4.38	4.21	0.17	0.278		
Age at first child's birth (years)						0.24	0.414
13-16	131	4.43	4.20	0.23	0.280		
17-19	153	4.28	4.29	-0.01	0.952		
Age of youngest child (years)						0.18	0.556
Under 1	106	4.26	4.06	0.20	0.384		
1 or older	178	4.39	4.37	0.02	0.895		
Number of pregnancies						0.29	0.394
1 or 2	213	4.44	4.27	0.17	0.320		
3 or more	70	4.06	4.18	-0.12	0.688		

(continued)

Table 4.11 (continued)

Characteristics and Subgroup at Baseline	n	Child's Affection to Mother		Within-Subgroup Impact	p[a]	Between-Subgroups Difference	p[b]
		Experimentals	Controls				
Education and Literacy							
Highest grade completed						0.14	0.641
10th or below	182	4.41	4.26	0.15	0.420		
11th or above	102	4.23	4.22	0.01	0.966		
Interval since last attended regular high school						0.44	0.137
2 years or less	130	4.48	4.14	0.34	0.117		
3 years or more	147	4.22	4.32	-0.10	0.626		
TABE reading test score (grade equivalent)[d]						--	0.497
7th grade or below	130	4.13	4.19	-0.06	0.788		
8th or 9th grade	78	4.53	4.22	0.31	0.257		
10th grade or above	76	4.57	4.31	0.26	0.352		
Employment and AFDC History							
Ever employed						0.25	0.506
Yes	221	4.39	4.23	0.16	0.356		
No	63	4.20	4.29	-0.09	0.793		
Family received AFDC when sample member was growing up						--	0.481
Never	82	4.52	4.34	0.18	0.539		
Sometimes	142	4.37	4.21	0.16	0.430		
Always	59	3.99	4.25	-0.26	0.404		
Psychosocial Characteristics							
CES-D (depression) score[e]						--	0.935
0-15 (not at risk)	137	4.36	4.25	0.11	0.604		
16-23 (at some risk)	76	4.36	4.33	0.03	0.915		
24-60 (at high risk)	71	4.30	4.12	0.18	0.546		
Self-Esteem score[f]						0.13	0.693
Below the mean	74	4.55	4.35	0.20	0.486		
At or above the mean (35)	210	4.27	4.20	0.07	0.677		
Number of sources of emotional support						0.07	0.824
0-2	127	4.35	4.21	0.14	0.522		
3 or more	157	4.34	4.27	0.07	0.722		

(continued)

Table 4.11 (continued)

Characteristics and Subgroup at Baseline	n	Child's Affection to Mother		Within-Subgroup Impact	p[a]	Between-Subgroups Difference	p[b]
		Experimentals	Controls				
Level of satisfaction with emotional support[g]						0.27	0.355
Less than very satisfied	135	4.55	4.29	0.26	0.224		
Very satisfied	149	4.17	4.18	-0.01	0.958		
Locus of Control score[h]						0.07	0.829
Below the mean	111	4.30	4.15	0.15	0.538		
At or above mean (22)	173	4.38	4.30	0.08	0.677		
Sample size		184	106				

Sources and Notes: See Table 4.8.

Table 4.12

Impacts of New Chance on Quality of Relationship, by Subgroups Formed at Baseline

Characteristics and Subgroup at Baseline	n	Quality of Relationship		Within-Subgroup Impact	p[a]	Between-Subgroups Difference	p[b]
		Experimentals	Controls				
Full Sample							
Mean		4.10	3.90	0.20	0.208		
Demographic Characteristics							
Child's age at observational study (months)						--	0.416
24-36	83	3.48	3.35	0.13	0.653		
37-47	114	4.22	3.81	0.41 *	0.098		
48-63	87	4.49	4.58	-0.09	0.764		
Gender of child[c]						0.26	0.416
Female	141	4.12	3.80	0.32	0.141		
Male	143	4.08	4.02	0.06	0.773		
Number of children						0.43	0.186
1	180	4.11	3.74	0.37 *	0.070		
2 or more	104	4.06	4.12	-0.06	0.812		
Race/ethnicity						0.30	0.562
Black	238	4.01	3.77	0.24	0.168		
White	46	4.59	4.65	-0.06	0.902		
Age of mother (years)						--	0.976
16-17	54	4.24	4.02	0.22	0.559		
18-19	146	3.99	3.83	0.16	0.467		
20-22	84	4.19	3.95	0.24	0.414		
Living arrangement (mother)						-0.25	0.446
Living with mother	102	4.11	4.07	0.04	0.875		
Not living with mother	175	4.06	3.77	0.29	0.159		
Living arrangement (partner/husband)						-0.86	0.131
Living with partner/husband	28	3.71	4.31	-0.60	0.273		
Not living with partner/husband	249	4.12	3.86	0.26	0.122		
Age at first child's birth (years)						0.05	0.857
13-16	131	4.10	3.87	0.23	0.317		
17-19	153	4.10	3.92	0.18	0.438		
Age of youngest child (years)						-0.14	0.654
Under 1	106	3.90	3.79	0.11	0.673		
1 or older	178	4.21	3.96	0.25	0.224		
Number of pregnancies						0.37	0.312
1 or 2	213	4.16	3.88	0.28	0.128		
3 or more	70	3.92	4.01	-0.09	0.782		

(continued)

Table 4.12 (continued)

Characteristics and Subgroup at Baseline	n	Quality of Relationship		Within-Subgroup Impact	p[a]	Between-Subgroups Difference	p[b]
		Experimentals	Controls				
Education and Literacy							
Highest grade completed						-0.14	0.692
10th or below	182	4.13	3.99	0.14	0.472		
11th or above	102	4.03	3.75	0.28	0.289		
Interval since last attended regular high school						0.21	0.521
2 years or less	130	4.17	3.88	0.29	0.215		
3 years or more	147	4.00	3.92	0.08	0.694		
TABE reading test score (grade equivalent)[d]						--	0.119
7th grade or below	130	3.90	4.04	-0.14	0.556		
8th or 9th grade	78	4.35	3.75	0.60 **	0.047		
10th grade or above	76	4.22	3.82	0.40	0.178		
Employment and AFDC History							
Ever employed						0.12	0.763
Yes	221	4.16	3.93	0.23	0.214		
No	63	3.89	3.78	0.11	0.749		
Family received AFDC when sample member was growing up						--	0.802
Never	82	4.06	4.02	0.04	0.906		
Sometimes	142	4.17	3.89	0.28	0.217		
Always	59	3.98	3.87	0.11	0.734		
Psychosocial Characteristics							
CES-D (depression) score[e]						--	0.732
0-15 (not at risk)	137	4.07	3.83	0.24	0.293		
16-23 (at some risk)	76	4.20	4.19	0.01	0.966		
24-60 (at high risk)	71	4.03	3.69	0.34	0.280		
Self-Esteem score[f]						0.07	0.848
Below mean	74	4.29	4.04	0.25	0.421		
At or above mean (35)	210	4.03	3.85	0.18	0.317		
Number of sources of emotional support						0.00	0.988
0-2	127	4.09	3.89	0.20	0.393		
3 or more	157	4.10	3.90	0.20	0.356		

(continued)

Table 4.12 (continued)

Characteristics and Subgroup at Baseline	n	Quality of Relationship		Within-Subgroup Impact	p[a]	Between-Subgroups Difference	p[b]
		Experimentals	Controls				
Level of satisfaction with emotional support[g]						0.15	0.613
Less than very satisfied	135	4.44	4.13	0.31	0.169		
Very satisfied	149	3.81	3.65	0.16	0.471		
Locus of Control score[h]						-0.25	0.445
Below mean	111	3.93	3.88	0.05	0.827		
At or above mean (22)	173	4.21	3.91	0.30	0.142		
Sample size		184	106				

Sources and Notes: See Table 4.8.

than others. The subgroups in which we were interested are the same as those listed in Table 4.5 — subgroups formed from baseline information about the mothers' lives. The rest of this section summarizes the subgroup findings for New Chance impacts on the mother, child, and relationship affective quality scales that yielded significant results.

Mother's Harsh Treatment. There were many findings for this scale, which is not surprising since it had significant overall program impact findings. The subgroup findings seem to fall into two general categories of positive program impact: lower Harsh Treatment scores for both low- and high-risk subgroups in New Chance, representing a broad effect for almost all subgroups. Two lower-risk subgroups who experienced positive New Chance impacts, for example, were mothers who felt a higher level of control over their lives and mothers who had been employed at some point in their lives. Two examples of higher-risk subgroups with positive program impacts were mothers who had less than an 11th-grade education or whose youngest child was less than one year old at baseline. Overall the findings suggest that the New Chance Program had a broad impact on reducing the amount of harsh treatment in the mother-child relationship.

Child's Compliance. Surprisingly, findings for this scale revealed that, for selected subgroups, children of mothers in the experimental group showed less compliance than children of mothers in the control group. Specifically, the subgroups were those in which the children were between 48 and 63 months old at the time of the observation, where the mother had more than one child at baseline, where the mother lived with a partner at baseline, and where the mother was at some risk for depression. We believe that these subgroups represent mothers who had many demands on their personal resources at baseline. The New Chance program may have introduced additional demands on these families, resulting in some disruption of mother-child relationships (e.g., an increased use of daycare), and consequently less compliance from the children.

Quality of Relationship. Dyads in the New Chance experimental group had higher scores than dyads in the control group for three of the baseline subgroups: where the child was between 37 and 47 months old at the time of the observation (the target age for the Teaching Tasks), where the mother had only one child at baseline, and where the mothers had an 8th- or 9th-grade TABE reading score. Being in one of these subgroups might indicate lower risk, and the finding may represent an intervention effect for the more advantaged subgroups. The New Chance Program seems to have increased the affective quality of the mother-child relationship for dyads in these subgroups.

Summary. By examining the impact of the New Chance Program on different subgroups (which were formed from questions asked at baseline), we can discern three general trends. For lower-risk subgroups, assignment to the New Chance experimental group seemed to increase the quality of the mother-child relationship in the Teaching Task session. For both lower-risk and higher-risk subgroups, participation in New Chance seemed to reduce mothers' harsh treatment of their children. For some mothers who already had high demands on their resources at baseline, the additional demands of the New Chance Program may have resulted in less compliance from their children.

IV. Conclusions

In this chapter we have described the affective quality coding scheme used to assess qualities of the interaction between mothers and their children as they carried out a set of Teaching Tasks. We have provided descriptive information about the affective quality rating scales. We have also related information about maternal and child characteristics at baseline and at 18 months after baseline to the qualities of the mother-child interaction on the Teaching Tasks. In addition, we have provided information about the impact of the New Chance Program on the affective quality of mother-child interaction in the Teaching Task session. Overall, we found few impacts of the New Chance Program, either positive or negative, on mother-child interactions.

In the analyses that looked at subgroup membership at baseline and at 18 months, the results suggest that some maternal characteristics predicted more positive interactions during the Teaching Task session. These characteristics, such as the mother feeling in control of her life and having fewer difficult life circumstances, suggest a common theme of attenuated maternal stress. We believe that this set of findings may indicate that mothers who are experiencing less stress in their own lives have more positive interactions with their children.

In an earlier presentation on some of the findings of the New Chance Observational Study (Egeland et al., March, 1995), it was noted that New Chance field representatives found it necessary to prioritize their goals with regard to parenting and the mother-child relationships. (For details on some of the challenges encountered by parenting educators, see Chapter 3 of this monograph.) The first and most difficult goal was to reduce the harsh, punitive parenting that is often seen in families living in poverty. The second goal was to encourage more positive parenting behaviors.

The major impact of New Chance on mother-child relationships was that mothers assigned to the New Chance experimental group, in comparison with those in the control group, showed less coercive and punitive behavior toward their children. This finding was true as an overall difference between the groups as well as within a wide range of subgroups. Thus, it seems as though New Chance did achieve the first goal — reducing harsh treatment of children by their mothers.

There were no overall differences between the New Chance experimental group and the control group on more positive indices of interaction such as Mother's Supportive Presence or Quality of Relationship, although the means for the New Chance experimental group mothers were higher than those for the control group mothers. We did see significant differences within some of the subgroups that may have been indices of lower risk at baseline, such as having only one child at the time of random assignment. Within these lower-risk subgroups, mothers in the experimental group were rated higher on Quality of Relationship than were control group mothers. Although this finding must be interpreted cautiously due to the lack of an overall New Chance impact, it does suggest that mothers who were at lower risk at baseline may have benefited from New Chance not only through the reduction of punitive and coercive parenting behaviors, but also through an increase in positive parenting behaviors. Thus, New Chance may have begun, for some enrollees, to achieve the second goal.

One of the major goals of the comprehensive New Chance Program was to have a positive impact on the children. We did not, however, find any overall impacts on observed child behavior in this evaluation. In fact, baseline subgroup analyses suggest that for some dyads, particularly those where the mothers faced interpersonal demands such as having multiple children living with partners, children of mothers in the New Chance experimental group showed less compliance than children of mothers in the control group. Again, subgroup findings must be interpreted cautiously due to the lack of an overall impact of the New Chance Program on Child's Compliance. This finding suggests that for some families the additional demands of the New Chance Program may have disrupted some aspects of the mother-child relationship. Previous research has suggested that multiple maternal stressors may contribute to child behavior problems in the preschool years (Richman, Stevenson, and Graham, 1982). For a small group of families, it is possible that the intensive nature of the New Chance Program compounded the demands in their already stressful lives.

The pattern reported here of positive maternal effects without positive child effects may have come about because New Chance, although a two-generational program, was still primarily mother-focused (see Monograph Introduction for a description of the program). Children were provided with daycare as well as enhanced access to health services, but these two direct child interventions may not have been enough to foster positive child outcomes in a short-term evaluation. In order to produce positive changes in child behavior over a short duration, more intensive, direct child components might be necessary. Improvements in the mothers' lives and parenting skills brought about by New Chance are important and noteworthy. Positive changes in the children's lives may come about only indirectly, however, and may depend on their mothers maintaining what they gained from New Chance.

Chapter 5

Mother-Child Interactions Related to the Emergence of Literacy

Jeanne De Temple and Catherine Snow

Children experienced in the use of extended language (that is, children who have experienced a rich language environment) at home in the preschool period are more prepared for both the language and the literacy demands of school. Mothers with more formal schooling and a higher level of literacy generally demonstrate greater skill with extended language and a greater tendency to use it with their children. Since the New Chance Program provided classes in education, literacy, and parenting, it would be expected to have an impact on the amount of extended language that mothers used with their children. Within the New Chance Observational Study, this expectation was assessed in two contexts: the type of talk mothers produced while reading a book with their children and the type of help mothers gave their children during a game involving thinking of objects with wheels.

Indeed, mothers assigned to the New Chance experimental group scored higher than controls on one global measure of performance during book reading — Book Reading Quality — though not on the amount of nonimmediate talk produced. Contrary to expectation, mothers in the experimental group scored lower than controls on Ease of Ideas (facility in carrying out the task) during the wheels task, a finding that was driven by those who were in more disadvantaged groups (for example, less education, lower literacy, greater risk for depression, lower sense of personal control).

Although the mothers in this study talked more during book reading than other low-income populations studied previously, nonimmediate utterances, the measure of extended language in this context, was quite rare. Mothers with more formal education and higher literacy scores at baseline and at the time of the observational session were rated higher on Book Reading Quality. Mothers of older children used more sophisticated language, suggesting they were adapting to the age of the child. Mothers more at risk for depression at the time of observation and those who were pregnant during this period scored lower on Book Reading Quality, as did mothers who had had less of a sense of personal control at baseline. The use of child care was associated with better book reading (higher Book Reading Quality, more complex discussion, and less talk). The wheels task was most successfully carried out by the older children and their mothers and by those mothers with more formal education and higher reading scores at the time of observation. The children with some child care were more successful in producing correct answers in this game.

I. The Importance of Mother's Talk with Her Child in Language Acquisition and in the Development of Discourse Skills Required in School

We report here on an analysis of the way mothers talk to their young children during two different interactive contexts — while reading a book together and while playing a guessing game. We focus on the talk of mothers in these contexts because we believe such talk gives us a window onto experiences that are preparing the children for participation in school and for the acquisition of literacy. Mother-child talk is, of course, a context in which children learn language. Much research has been devoted to parent-child interaction as it relates to child language acquisition (see Gallaway and Richards, 1994). More specific to our purposes, though, mother-child interaction is a context in which children can gain specific language skills — skills in the kind of discourse that predominates in academic and literacy-related settings as early as kindergarten.

We start from the assumption that language can be used in many different ways, some of which are particularly important for academic success. In particular, children who will do well in school seem to be skilled in "distanced" communication, or talking in ways that are communicatively effective to a distant or unknown listener. This type of talk, referred to as decontextualized language, is characterized by such key features as the speaker's awareness of the importance of filling in nonshared background knowledge, providing new information systematically, and referring to people or objects using explicit lexical items rather than pronouns (it, this). Children skilled in the use of decontextualized language are good at making connections between new information and information they already have, relating present events to past ones, and associating current puzzles with previous solutions. Such children are typically capable of organizing a narrative so that important information is distinguished from background information, providing explanations and definitions, and producing stretches of talk with minimal conversational support.

A. Mothers' Use of Decontextualized Language

Decontextualized language is of particular interest because decontextualized language skills enable children to communicate effectively to a distant interlocutor and about abstract or nonpresent topics. Such skills constitute a context for obtaining abstract knowledge (for example, knowledge about history, science, literature) and a context for analyzing such knowledge. They are crucial in the acquisition of information because much of the knowledge required to function well in school is abstract, theoretical, and related to unknown people and past events. Even talking about such knowledge requires decontextualized language skills, and certainly acquiring such knowledge from texts or conversation does as well.

In the New Chance Observational Study we consider mothers' use of decontextualized language or strategies to encourage analysis of information in two contexts: book reading and the discussion thereafter, and a guessing game. For example, the discussion of *The Very Hungry Caterpillar* provides a context to talk about such topics as nutrition, digestion, and metamorphosis. The guessing game provides a context in which a fairly accessible body of knowledge about objects with wheels is made difficult by virtue of the guessing game format. Mothers have the opportunity to model strategies for thinking about and analyzing this body of knowledge in interaction with their children. Because children's skills in analyzing information are central to their school success, we expect that differences among mothers in their effectiveness in carrying out this task (assuming that this is indicative of the quality of their other interactions with their children) might well influence children's school outcomes and cognitive abilities.

B. Book Reading as a Context for Mother-Child Interaction

Book reading with young children is a widespread practice in literate cultures and households, though the specifics of how it is carried out certainly differ as a function of culture, group, and family. Book reading with children is widely cited as a practice that will promote child development and school success; the Family Literacy Center, Reading Is Fundamental, the Association of American Librarians, and many researchers (for example, Goldfield and Snow, 1984; Heath, 1983; Wells, 1985) all agree on its value. It is clear from the research on parent-child book reading, though, that the label covers a wide variety of activities, from looking at pictures and talking about them, or parent reading the text, to child reading aloud to parent and actual attempts to teach children to read. Most previous research has focused on quantity of book reading as a predictor of child outcomes; here we shift the focus to the quality of the interaction that occurs during book reading.

Differences in style of book reading become particularly important when we consider book reading as preparation for early school experiences, where teachers typically read books to children at circle time or meeting time. Teachers read books to small groups of children in a variety of different styles — for example, straight text reading with discussion afterwards, reading interspersed with questions and discussion, a dramatic performance style, and reading with echoing from the children. Teale and Martinez (1986), for example, document a difference in emphasis between teachers who focus on the books as stories versus print and labeling. Reading aloud is a regular part of daily group time in preschools and continues to be an important element of the curriculum in kindergartens and first grades. In O'Rourke's analysis of reading aloud in kindergarten through grade 6 (1990), teachers reported that they read aloud in order to improve children's oral and written language. Teachers also reported encouraging discussion of the books at every grade level as an additional support to language development; the most frequent topic of discussion reported was the relevance of the content of the story to the students' lives. Teachers' efforts to develop children's comprehension skills were most often coordinated with the materials chosen for reading aloud in the classroom. The teachers were not using reading aloud to focus predominantly on print, labeling, or recall of factual information, but rather as an opportunity for extended discourse, for providing and listening to explanations, inferences, and connections to real-world experiences.

In a longitudinal study of book reading at home in low-income families (De Temple, 1994), mothers' comments and questions while reading to preschool-aged children were coded as "immediate" talk (for example, labeling, counting, paraphrasing) or "nonimmediate" talk (for example, explanations, inferences, predictions, real-world connections). Mothers' use of nonimmediate talk when the children were 3½ years old was strongly associated with children's scores in kindergarten on measures that are good predictors of later reading skill: oral story comprehension, receptive vocabulary, definitions (giving clear explanations of the meaning of simple words), superordinates (knowing how to name categories such as furniture, vehicles, tools), and emergent literacy (prereading skills such as knowing how to handle a book, recognizing common signs, knowing letter names). The percentage of immediate talk used was negatively associated with these language measures.

It is hard to know exactly what aspect of book reading is the most influential in children's development; reading books provides experience with print, exposure to new vocabulary, and practice with complex language structures. Furthermore, homes where book reading occurs tend to be homes where interesting talk goes on in other contexts as well and where additional activities that are supportive of literacy development occur. Thus, book reading quantity and quality may themselves be correlated with variables that promote development. Book reading can undoubtedly be a context for the acquisition of print skills (letter recognition, simple word reading) and for growth in children's understanding of literacy as a social and cognitive resource. We argue, though, that perhaps the most powerful effect of book reading on the development of young children comes through the talk that goes on in the context of book reading rather than through effects on literacy skills narrowly conceived. Our focus in the New Chance Observational Study will be on the level of talk in the context of book reading and the discussion following it.

C. The Wheels Task

The wheels task was developed by Byron Egeland and his collaborators (see Chapter 4) as one of their series of challenging interactive tasks. While it has been studied previously as a context for observing characteristics of mother-child interaction globally, our use of it here is the first attempt to analyze the linguistic details of the wheels task interactions, coding maternal speech for organizational features, cognitive support, and feedback.

The wheels task is of interest because it provides an opportunity to observe mothers' strategies for identifying a category or superordinate-level concept in a way adapted to their children's developmental levels. Children aged 2 or 3 are able to categorize (for example, name "things to wear" or "things to eat"), but they do not always have full verbal access to the higher categories in their taxonomies, as shown by the difficulty they have in answering questions like "what do you call a pear and an apple?" that require superordinates ("fruit") as the response.

The wheels task is, in a way, the inverse of the task of defining a word, which we have found is a good indicator of children's decontextualized language skills. A child asked to define the word "car" might be expected to say something like "a car is a machine with wheels and a motor that you can ride in," relating "car" to the general category of "things with wheels" as well as differentiating it from other members of that category. Since we have found that giving definitions that specify category membership is a correlate of literacy and school achievement for young children (Snow et al., 1995), seeing how mothers model talk about categories and category members is of considerable interest.

D. Examining These Particular Behaviors and Contexts as Part of the New Chance Intervention and Evaluation

Previous research has shown that both the quantity (Wells, 1985) and quality (Whitehurst et al., 1988) of book reading occurring at home predict children's later language and literacy performance and that quality and quantity measures predict language and literacy achievement differentially (De Temple, 1994). The observational study described here provides the opportunity to examine the *quality* of book reading in the home and the impact of the New Chance Intervention on this fine-grained measure of linguistic and cognitive stimulation in mother-child interaction.

Two possible pathways through which New Chance might affect a mother's book reading behavior with her young child include changing the mother's behavior as a learner and as a teacher of her child. The first pathway identifies as crucial the mother's own education and development, which might be affected by participation in formal education. This in turn could affect the quality of the mother's reading by providing her with strategies for learning in general, and reading in particular, that improve comprehension, explanations, and inferential thinking. The second pathway identifies as crucial development of the mother as a teacher, with improved knowledge of and attention to her child's needs. The New Chance parenting education classes may have provided discussions of cognitively stimulating and developmentally appropriate interactions with children and the value of book reading. In addition, a mother's observation of her child's child care program may have provided opportunities for learning about both children's skills and engaging children in book reading.

For many of the same reasons, we would expect participation in New Chance to have an impact on both mothers' strategies for eliciting the names of objects with wheels and children's success in the guessing game. Previous research has demonstrated that the ability to provide formal definitions is associated with years of formal education (Tapia-Uribe, 1988) and school success, particularly in literacy (Snow et al., 1989; Watson, 1989). Thus, mothers' own enhanced education through New Chance could enhance their ability in a context involving definitions. Parenting education programs may help mothers understand their children's developmental stages, thereby enhancing their ability to choose appropriate strategies for their children in the guessing game. Children attending child care programs may also have experienced this type of verbal game and respond more readily to their mother's questions.

In the remainder of the chapter we examine the two tasks in the New Chance Observational Study that are particularly important to emergent literacy: reading a book together and playing the guessing game "What has wheels?" We first describe the coding systems applied to each task and then present descriptive analyses of the mothers' behavior in each context, examining the relationship between measures of maternal behavior and characteristics of the mothers at the time of their enrollment in the study (baseline). We then examine the relationship between performance on these tasks and characteristics of the mothers and children 18 months later (shortly before the tasks were carried out). Finally, we examine the evidence for program impacts on mother-child interaction during each task, both overall and for key subgroups.

II. The Coding of Literacy-Related Interactions in the New Chance Observational Study

A. The Coding of Behavior in Book Reading and Book Discussion

Each book reading session (including both reading and discussion) was transcribed from the videotapes into computer files in the transcription format required for analysis by the software available through the Child Language Data Exchange System (CHILDES; MacWhinney and Snow, 1985). Coding was carried out utterance by utterance for the mother and her child.[1] Every utterance by the mother and child during the reading of the book (excluding the reading of the text of the book) and during the discussion of the book received a code. Additionally, the interaction during book reading was globally coded on three measures: the mother's Comfort Level, Reading Intonation and Animation, and Reading Fluency.

1. **The Coding of Book Reading.** One focus of the coding of talk *during* book reading addresses the extent to which the book reading interaction provides opportunities for different types of language skills by distinguishing between *immediate* talk, which uses information that is readily available from the illustrations or text (such as labeling, counting, naming colors, paraphrasing the text, requesting immediate recall of the text), and *nonimmediate* talk, which requires going beyond the text or illustrations to make predictions and connections to the real world, draw inferences, analyze information, discuss the meaning of words, and give explanations.

The following excerpts from transcripts of a mother reading *The Very Hungry Caterpillar* provide examples of immediate and nonimmediate talk. In the first example, the mother reads to Tyrone (47 months old) focusing on factual information depicted in the illustrations and thus using immediate talk.[2]

Mother: "One Sunday morning the warm sun came up and *pop* out of the egg came a *tiny* and a very large caterpillar." Look Tyrone! See the sun? (Points to sun.)

Tyrone: (Nods.)

Mother: What's this? (Points to caterpillar.)

Tyrone: I *saw* that. (Points to caterpillar)!

Mother: Caterpillar.

Tyrone: Caterpillar.

Mother: "He started to look for some food."

---

[1] An utterance is defined as the unit of speech planning, that is, a speech sequence with a complete intonation contour, followed by a pause or speech by the conversational partner. Utterances typically also represent completed syntactic units, for example, sentences, clauses, or phrases, though interjections (wow!) or responses (mm-hmm, huh?) also count as utterances.

[2] The names of the children in all of the transcripts have been changed to protect their privacy.

— —

Mother: "On Friday he ate through *five* oranges but he was still hungry." How many?

Tyrone: Two one three four.

Mother: How many? Start here!

Tyrone: One two three four five. (Mother points to oranges as Tyrone counts.)

Mother: Right.

In the second example, Keisha (age 45 months) and her mother use nonimmediate talk when they make predictions, ask for inferences, and provide explanations.

Mother: "... One cupcake and one slice of watermelon."

Keisha: One watermelon.

Mother: And guess what? Later that night do you know what happened to him? The caterpillar? "That night he had a stomachache." Eating all that food that's why.

Keisha: Eating all that food!

Mother: See? See how he's looking so sad and pitiful? Guess what Keisha? "The next day was Sunday again. The caterpillar ate through one nice green leaf. He ate through one nice green leaf and after that he felt much better." Why'd he feel much better I wonder. Hmmm, let's find out.

Keisha: Let's find out.

Mother: Okay.

While an extensive coding system was applied, here we focus on the summary variables of Number of Nonimmediate Utterances, Percentage of Immediate Utterances, and Total Number of Utterances. Only those utterances that occurred *during* the book reading, as defined below, were coded with this scheme. Coding started with the reading of the title of the book or, if the title was not read, with the first utterance after the first line of text was read and continued through one utterance after the last line of text had been read, or until the mother indicated that the activity was over (for example, "the end"). This decision was made in order to define clearly and consistently the context to be sampled (a book reading session).

Although all mother and child talk was coded, our analyses will focus only on the mother's talk during book reading. The book reading context is one in which the reader directs the activity and initiates most of the interaction that occurs. Furthermore, the questions being asked in this study focus on the ways in which the mother uses the book reading context as an opportunity for extended discourse. Therefore, the focus of the analysis is on maternal book reading variables.

2. **The Coding of the Book Discussion.** The coding of the discussion focused on the occurrence of specific topics. The decision was made simply to count the number of different

topics included in any discussion, rather than, for example, total length of discussion, because topic variety appeared to be the best marker of sophistication, complexity, and interest level of the discussion. Some relatively long discussions were quite boring, for instance if they consisted entirely of a review of the names of objects pictured.

The following list of possible discussion topics was generated by examining all utterances by the mothers and children during the discussion that were relevant to the book:

1. Picture prompt: Mother pages through the book asking the child to name objects.
2. Conclusion: The fact that the caterpillar turns into a butterfly is mentioned.
3. Evaluation: An evaluation is made or requested (for example, Did you like that?).
4. Real world: Some connection is made to the child's experience or to the real world generally.
5. Summary: Mother summarizes the text without engaging the child.
6. Memory: Mother asks the child to recall facts from the story with the book closed.
7. Request recap: Mother asks the child to summarize the story but with the book open.

Scores on topic types ranged from zero, for dyads in which no book-related discussion occurred, to seven, for dyads whose discussions touched on every topic. This score reflects the scope and complexity of the book reading discussion.

3. The Global Coding of Reading Interaction. The global ratings for book reading were included to capture an element of the book reading experience that is not necessarily reflected in the utterance coding. We included global ratings because readers who have very little conversation about the book may nonetheless successfully engage children through an effective, animated, lively reading style that displays their own enjoyment and comprehension. Alternately, a halting, awkward reading style may impede the child's comprehension or interest. Coders rated each mother on a 3-point scale for each of the following dimensions: Reading Intonation and Animation, Comfort Level, and Reading Fluency. A summary Book Reading Quality measure was derived by summing these three maternal variables. Book Reading Quality could range from a low score of 3 to a high score of 9.

4. A Summary of Variables Based on Book Reading and Discussion. For each book reading session the following five variables were examined. One goal in selecting variables was to identify statistically independent measures that reflect both the amount and the cognitive quality of the interactions.

- Total Number of Utterances: The total number of utterances by the mother during the book reading, a reflection of the amount of interaction. This includes all utterances that occurred within the unit of analysis, including those that were uncodable owing to audio difficulty or unintelligibility and those that were irrelevant to the activity.

- Number of Nonimmediate Utterances: The sum of all maternal utterances, questions, or comments, coded as nonimmediate (for example, predictions, text-reader links, inferences, explanations). This measure reflects the amount of exposure the child has during each session to cognitively challenging talk.
- Percentage of Immediate Utterances: The percentage of maternal utterances that are coded as immediate (for example, color naming, labeling, counting, paraphrasing of the text). This variable was obtained by dividing the number of maternal immediate utterances by the total number of maternal utterances. It reflects the mother's relative preference for less challenging yet book-related talk.
- Number of Discussion Topics: The sum of the discussion topic types that occurred during the discussion.
- Book Reading Quality: The sum of the ratings of maternal book reading: Reading Fluency, Reading Intonation and Animation, and Comfort Level.

5. Reliability. The method of coding mother's and child's talk during book reading had been applied in previous research involving mothers and preschool-age children from low-income families, using a variety of books including *The Very Hungry Caterpillar* (De Temple and Beals, 1991; Dickinson et al., 1992; De Temple, 1994). In this earlier research, reliability between coders was very high. A Cohen's Kappa measure of reliability (Landis and Koch, 1977) ranged from .82 to .88 for the various books. Reliability estimates above .81 are considered "almost perfect," suggesting that this coding scheme can be learned with minimal variation in interpretation among coders.

Two coders were trained to code the New Chance Observational Study transcripts. Following training, the first 10 percent of each coder's transcripts was coded in collaboration with the trainer. In addition, any confusion in further coding was resolved in conference between the coders. The Cohen's Kappa measure of reliability for another 10 percent randomly chosen from the 290 coded transcripts of the New Chance Observational Study was .85 on the immediate/nonimmediate dimension. Total Number of Utterances is generated automatically by the software package (CLAN) used in the analysis.

The discussion coding involved noting the presence or absence of specific topics and could be coded with complete reliability. Global ratings were assigned by the same researcher who transcribed the book reading. All global ratings were verified by a single experienced transcriber and coder. Any discrepancies were discussed with a third coder and a consensus was reached.

B. The Coding of Behavior During the Wheels Task

The mother-child interaction during the wheels task was transcribed from videotape into computer files in the format needed for analysis by the software available through the Child

Language Data Exchange System (CHILDES; MacWhinney and Snow, 1985). Coding was carried out utterance by utterance for the mother and her child. Additionally, the coder rated the mother on a global measure, Mother's Ease of Ideas, in carrying out the task.

1. **Variables Based on the Wheels Task**. An extensive coding scheme was applied to the transcripts in order to analyze maternal strategies and child responses during the wheels task. For the purposes of this analysis the codes have been summarized to reflect the child's success in responding to the task (that is, the number of wheeled objects that the child named) and the number of elicitations the mother produced to obtain each correct response from her child. Some mothers had to question their children several times and use a variety of prompts to assist them in naming a wheeled object; this often occurred when mothers used relatively obscure hints, produced unclear delimitations of the target category, or simply repeated previously unsuccessful requests. Other mothers, evidently more tuned in to their children's knowledge, were able to ask a direct question or provide one clue to elicit the intended response. The number of wheeled objects named correctly by the child divided by the number of elicitations by the mother therefore provides a measure of the mother's overall effectiveness.

Each mother was rated on a scale labeled Ease of Ideas in the wheels task. This 4-point scale was used to measure skill in carrying out this task effectively and comprehension of appropriate strategies. A low score of 1 indicated confusion about the task leading to an inability to carry it out successfully (for example, asking the child to say "wheels"). A high score of 4 was given to a mother who clearly understood the game, had a variety of wheeled objects in mind, and effectively prompted her child, modifying her strategy if necessary in response to the child's reactions.

The following excerpts of transcripts of mothers and children carrying out the wheels task represent a range of strategies and effectiveness. In the first example, the mother successfully elicits the names of wheeled objects from her 52-month-old son, Michael. The mother makes sure the child understands the topic, orients her child to the task, asks explicit questions providing meaningful clues, and frequently refocuses the task with variations of "that has wheels."

Mother: You know what wheels are, right?

Michael: (Nods.)

Mother: What?

Michael: Um, they work like this. (Moves fingers along table.)

Mother: But what do the wheels be on? ... What do you — what do Michelle ride you in that bring you home?

Michael: Um, a car!

Mother: Okay. That got wheels. (Draws a circle in the air to demonstrate wheels.) Tires, they just like wheels.

Michael: Yeah. Okay.

Mother: Okay, what else got wheels like that?… What do you want to ride on? What do you want for your birthday?

Michael: A Jeep.

Mother: That got wheels, right?

Michael: Yeah.

Mother: Okay. What else? What do you be riding on?

Michael: Hmm, a bike.

Mother: That got wheels, right?

Michael: Yeah.

Mother: What else?

Michael: Um, oh, talk about Ninja Turtle toys.

Mother: Mmhm.

Michael: Because I know they got cars.

The mother in the following example uses very similar effective strategies with her 31-month-old daughter, but she needs to provide more prompts and support to elicit a correct response from a very young child.

Mother: Tara, you know what wheels are?

Tara: I wanna play with this.

Mother: You see wheels in here? There's two. What does that do? What is that over there? (Points to something across the room.) What is that? What is it?

Tara: A wheel.

Mother: It is! What's that that he rides?

Tara: On a bike.

Mother: On a bike? And there's wheels on the bike? You know what you drive? Look at Mommy. Look at Mommy. You know what Carmela drives? What does she drive?

Tara: A car.

Mother: Oh does that have wheels?

Tara: Cars. (Nods.)

Mother: Cars do. And what else? What does Tonio, what does he play with?

Tara: A toy.

Mother: Okay. What kind of toy? Look at Mommy. Look at Mommy. What kinds of toy? What is that big yellow toy that he plays with? It's got wheels on it. Huh? You know what it is. You played with it yesterday. It has wheels on it. What is it? And it go beep beep beep.

Tara: Beep.

Mother: What is it? … okay the wheels on the what go round and round? (Makes slight round movement with hands.)

Tara: Round and round.

Mother: What's that song? What is it?

Tara: Wheels on it.

Mother: What does Jerry drive?

Tara: The wheels on the bus.

Mother: He drives a what?

Tara: The wheels on the bus.

By contrast, the mothers in the following examples did not carry out the wheels task effectively. Angela, 43 months old, was not given an opportunity to name objects with wheels. After the first question, her mother simply asks her to repeat words. While she is correct in her statement that Angela won't "just come out and name them," she does not attempt to assist her in generating her own list of objects with wheels.

Mother: What got wheels?

Angela: Hmm?

Mother: Don't your bike got wheels?

Angela: (Nods.)

Mother: She know they do. She ain't going to just come out and name them. What else got wheels? … A car.

Angela: (Nods.)

Mother: Say a car.

Angela: A car.

Mother: A bike.

Angela: A bike.

Mother: What else? Motorcycle.

Angela: Motorcycle.

Mother: Trucks.

Angela: Trucks.

Mother: Hmm? Skateboard.

Angela: Skateboard.

In the following example, the mother receives a rating of 1 for Ease of Ideas because she has completely misunderstood the task and attempts to elicit the word "wheels" from 31-month-old Jason.

Mother: Oh, um, um, I don't know how. Um, don't have nothing to show him. … What makes your truck move?

Jason: (Points toward the truck.)

Mother: Yeah, what makes it move? What are those things down there? Up under it?

Jason: (Unintelligible response.)

Mother: What's on a bike?

Jason: Hmm?

Mother: What's on a bike?

Jason: Bike?

Mother: Bike. Yeah, your bike's up there. What makes your bike move?

Jason: Huh?

Mother: What makes your bike move? What make it go? What are those things at the bottom? (Makes a circular motion with her finger.) Hmm? What makes your wheels move? Oh whoops!

The following three variables will be examined for the wheels task:

1. Objects Named: number of wheeled objects named correctly by the child.

2. Objects/Elicitations: wheeled objects named correctly as a proportion of maternal elicitations (a score below 1.0 indicates more than one elicitation per correct response).

3. Mother's Ease of Ideas: global rating of the mother's comprehension of and facility in carrying out the task.

2. Reliability. Two coders were trained to code the transcripts. Following training, the initial 10 percent of each coder's transcripts was coded in collaboration with the trainer. In addition, any confusion in further coding was resolved in conference between the coders. The Cohen's Kappa estimates for child's naming of wheeled objects calculated on a second 10 percent of the transcripts was .85 and for mother's elicitations was .84.

The global rating (Ease of Ideas) was assigned by one coder and verified by a second experienced transcriber and coder. Any discrepancies were discussed with a third coder and a consensus was reached.

III. Findings

A. Book Reading and Discussion

1. Descriptive Analyses. Mothers generally read the book to their children, asked questions and commented while reading, and discussed the book after completing the reading. However, 36 mothers (12.4 percent) did not read the text. There is no evidence that these mothers were not able to read the text. The literacy scores for these mothers on average did not differ from those who did read aloud, and the lowest scores on the literacy test for this group were not the lowest scores for the sample. Mothers with lower literacy scores than any of these mothers *did* read the book to their children. Mothers who did not read the text did page through the book, telling the story, commenting, and questioning in a similar way to mothers who did read the text. The utterances of these 36 mothers were coded and are included in the analyses of the book reading variables: Total Number of Utterances, Number of Nonimmediate Utterances, Percent of Immediate Utterances, and Number of Discussion Topics. However, because they did not read, these mothers could not be included in the global rating of Book Reading Quality.

Mothers and children generally engaged in quite a bit of talk during the book reading, as can be seen in Table 5.1, which presents means, standard deviations, and ranges for all five book reading variables. The Total Number of Utterances, excluding any reading of the text, ranged from 1 to 195, with a mean of 50.76 utterances. Nonimmediate talk during book reading was rare; the mean Number of Nonimmediate Utterances by the mothers was 1.94, with a range of zero to 17. This type of talk constituted, on average, only 3.45 percent of the mothers' total talk (actual range = 0-25.71 percent). Most of the mothers' talk during book reading was immediate; the Number of Immediate Utterances by mothers ranged from 1 to 150, with a mean of 34.18.

Table 5.1

Means on Maternal Book Reading Task Variables, by Subgroups Formed at Baseline

Subgroup Formed at Baseline	n	Total Number of Utterances	Number of Non-immediate Utterances	Percent of Immediate Utterances	Number of Discussion Topics	Book Reading Quality
Full Sample						
Mean		50.76	1.94	68.30	2.56	6.54
Standard deviation		36.45	2.64	12.87	1.42	1.48
Range		1-195	0-17	18.50-100	0-7	3-9
Demographic Characteristics						
Gender of child[a]						
Female	142	45.95	1.94	69.12	2.54	6.66
Male	148	50.98	1.97	67.40	2.71	6.43
Number of children						
1	183	48.23	1.89	68.80	2.58	6.48
2 or more	107	49.00	2.06	67.36	2.70	6.64
Race/ethnicity						
Black	244	49.96	1.96	68.81 *	2.60	6.49
White	46	40.41	1.93	65.02	2.76	6.85
Age of mother (years)						
16-17	56	57.99 *	2.19	70.51	2.56	6.26
18-19	148	44.58	1.67	67.45	2.63	6.54
20-22	86	50.01	2.33	68.32	2.66	6.72
Living arrangement (mother)						
Living with mother	104	42.28 **	1.58 *	69.35	2.59	6.39
Not living with mother	179	52.13	2.19	67.60	2.63	6.62
Living arrangement (partner/husband)						
Living with partner/husband	30	53.84	1.61	63.16 **	3.02	6.46
Not living with partner/husband	253	48.12	2.02	68.78	2.58	6.55

(continued)

BEST COPY AVAILABLE

Table 5.1 (continued)

Subgroup Formed at Baseline	n	Total Number of Utterances	Number of Non-immediate Utterances	Percent of Immediate Utterances	Number of Discussion Topics	Book Reading Quality
Age at first child's birth (years)						
13-16	135	48.47	1.83	67.67	2.73	6.54
17-19	155	48.58	2.08	68.77	2.53	6.55
Age of youngest child (years)						
Under 1	107	45.41	1.97	69.77	2.38 **	6.55
1 or older	183	50.33	1.95	67.35	2.77	6.54
Number of pregnancies						
1 or 2	218	47.98	2.01	68.07	2.57	6.45 *
3 or more	71	49.78	1.77	68.27	2.81	6.81
Education and Literacy						
Highest grade completed						
10th or below	185	48.70	1.98	68.48	2.59	6.42 *
11th or above	105	48.24	1.92	67.84	2.69	6.76
Interval since last attended regular high school						
2 years or less	134	49.25	2.23 *	68.10	2.51	6.44
3 years or more	149	48.31	1.66	68.96	2.78	6.65
TABE reading test score (grade equivalent)[b]						
7th grade or below	132	49.47	1.77	67.92	2.56	6.07 ***
8th or 9th grade	79	49.69	2.39	68.00	2.72	6.82
10th grade or above	79	46.08	1.84	68.94	2.64	7.00
Employment and AFDC History						
Ever employed						
Yes	224	48.79	1.90	67.87	2.68	6.57
No	66	47.58	2.17	69.59	2.45	6.43

(continued)

Table 5.1 (continued)

Subgroup Formed at Baseline	n	Total Number of Utterances	Number of Non-immediate Utterances	Percent of Immediate Utterances	Number of Discussion Topics	Book Reading Quality
Family received AFDC when sample member was growing up						
Never	84	43.64	2.17	68.67	2.61	6.80
Sometimes	145	49.84	1.78	67.96	2.65	6.40
Always	60	52.08	2.06	68.29	2.59	6.52
Psychosocial Characteristics						
CES-D (depression) score[c]						
0-15 (not at risk)	141	48.23	1.85	69.16	2.56 ***	6.70
16-23 (at some risk)	77	51.47	2.16	68.79	2.94	6.50
24-60 (at high risk)	72	45.80	1.95	65.75	2.42	6.27
Self-Esteem score[d]						
Below mean	75	44.68	2.02	69.52 ***	2.62	6.63
At or above mean (35)	215	49.80	1.75	64.35	2.65	6.28
Number of sources of emotional support						
0-2	133	48.00	2.10	68.64	2.63	6.62
3 or more	157	48.95	1.84	67.92	2.62	6.48
Level of satisfaction with emotional support[e]						
Less than very satisfied	138	48.94	1.96	67.27	2.61	6.65
Very satisfied	152	48.16	1.95	69.11	2.64	6.45
Locus of Control score[f]						
Below mean	115	48.64	1.64	66.73	2.65	6.20 ***
At or above mean (22)	175	48.46	2.15	69.17	2.61	6.76
Sample size	290					

(continued)

Table 5.1 (continued)

Sources: New Chance baseline enrollment data and coded observational study variables.

Notes: Means are adjusted for experimental/control status in order to be descriptive of the overall sample. Statistical significance levels are indicated as *** < = 1 percent, ** < = 5 percent, and * < = 10 percent, which refers to the probability that sample means are different from each other only because of random error.

For subgroups that have more than two levels, the placement of asterisks does not indicate where the significant difference lies, but only that there was a significant difference among the set of means.

Calculations for this table used data for all 290 respondents for whom there were observational data coded for literacy, and 18-month follow-up survey data, including those with values of zero for outcomes and New Chance enrollees (i.e., experimentals) who did not participate in the program. The sample size may fall slightly short of the number reported because of missing or unusable items from some respondents' questionnaires and videotape problems.

[a] When a sample member had more than one child, her response refers to the randomly selected focal child.

[b] The test used to measure reading ability was the reading part of the Tests of Adult Basic Education (TABE). Most sites administered the survey form of the test, but some administered the full reading test.

[c] The Center for Epidemiological Studies Depression (CES-D) Scale is a widely used measure of depression; scores can range from 0 to 60.

[d] The measure of self-esteem used was the Rosenberg Self-Esteem Scale, a 10-item scale that assesses a person's global sense of self-worth. Scores can range from 10 to 50; 30 is considered the neutral midpoint.

[e] Enrollees were asked about their degree of satisfaction with the emotional support ("people who listen to you, reassure you, and show you they care") they received. Levels range from 1 (very dissatisfied) to 5 (very satisfied); however, because so few individuals indicated that they were less than very satisfied with their emotional support, levels 1 through 4 were collapsed into one category labeled "less than very satisfied."

[f] The Locus of Control Scale is a 6-item adaptation of the longer scale originally developed by Julien Rotter (1966). Scores can range from 6 to 30; 18 is considered the neutral midpoint.

BEST COPY AVAILABLE

This type of talk accounted for, on average, 68.30 percent of the talk during book reading (actual range = 18.50-100 percent).[3]

Mothers and children typically incorporated several topics in their discussions; the mean Number of Discussion Topics was 2.56, with a standard deviation of 1.42. The range of 0 to 7 indicates that some did not engage in a discussion at all, while others included every topic in a lengthy, varied interaction.

Book Reading Quality scores reflected the possible range (3 to 9), with an average of 6.54 (standard deviation = 1.48). Only 29 of the 290 mothers were rated 1 (low) on Comfort Level, indicating that in general mothers experienced little discomfort during the book reading session. On Reading Fluency, and Reading Intonation and Animation, on the other hand, a higher proportion of the mothers scored low (45 out of 254) and a lower proportion scored high than on Comfort Level (see Table 5.2).

2. Comparisons with Other Research. Although the codes used here were based on those developed by De Temple (1994) and applied in the Home-School Study of Language and Literacy Development (HSSLLD), a longitudinal study of low-income mothers and their children, the coding was adapted slightly because of observed differences in the interactions. In the New Chance Observational Study sample we observed a recitation style not seen in HSSLLD, in which mothers requested that children repeat literally the modeled words or phrase from the text. On the other hand, two types of utterances used by mothers and children in the HSSLLD sample were never observed in the New Chance Observational Study sample: text-text links (comments or questions relating the book being read to other books known to the child) and print-related talk (for example, asking the child to name letters or to sound out words). It is striking that although the New Chance Observational Study sample included children as old as 5, neither of these two literacy-focused uses of the book reading interaction occurred.

Mothers in the New Chance Observational Study produced much more talk during book reading than the mothers in the HSSLLD study (50.76 versus 38.91 Total Number of Utterances). Despite the greater amount of talk by New Chance mothers, they used fewer Nonimmediate Utterances (mean = 1.94) than the HSSLLD mothers (mean = 3.15). This type of talk constituted, on average, only 3.45 percent of the New Chance mothers' total talk versus 10.5 percent for the HSSLLD mothers. Immediate Utterances represented a higher percentage of the total talk for New Chance mothers (68.30 percent) than for HSSLLD mothers (61.25 percent). Since we have found that maternal use of nonimmediate talk relates to later child outcomes, in particular to their language and literacy skills (De Temple, 1994), the very low proportion of nonimmediate talk produced by the New Chance mothers is troubling. The HSSLLD sample, while comparable to the New Chance sample in child age and maternal education, included only 50 percent of single mothers on welfare and only a few mothers who had started their families in adolescence. The HSSLLD families, furthermore, volunteered to participate in a longitudinal research study that offered them no services or intervention. Volunteering for such a study may

---

[3] A small proportion of maternal talk consisted of organizational or irrelevant comments, not discussed further here. Maternal feedback to children's responses was not coded as either immediate or nonimmediate, and thus is also not included. Although feedback to responses is an important area of research, it does not directly address cognitive and linguistic complexity, the areas of focus for this analysis.

Table 5.2

Global Rating Frequencies for Book Reading Quality

	n	Low (Rating of 1)	Middle (Rating of 2)	High (Rating of 3)
Comfort Level	290	29(10.3%)	153(52.7%)	108(37.0%)
Reading Fluency	254	45(15.4%)	119(40.8%)	90(31.2%)
Reading Intonation and Animation	254	45(15.4%)	148(51%)	61(20.9%)

Source: New Chance coded observational study variables.

Note: The 36 mothers who did not read the text could not be rated for fluency or intonation and animation of reading.

indicate a somewhat higher level of interest in promoting children's school achievement.

3. **Baseline Characteristics and Book Reading Performance.** In this section, we consider the relationship between background variables and book reading performance, seeking possible differences in book reading between groups of mothers distinguished from one another on demographic, educational, and psychosocial characteristics that were present at baseline (see Table 5.1). In general, we discuss only findings that hold for at least two variables reflecting book reading performance, though all significant findings for Book Reading Quality (a summary variable) are discussed.

Demographic Characteristics: Mothers not living with their own mothers at the time of enrollment in the New Chance Evaluation had a higher Total Number of Utterances and Number of Nonimmediate Utterances than those living with their own mothers. Relatively more talk is associated with book reading to younger children. The higher Number of Nonimmediate Utterances, although usually associated with talk to older children, may in this case be tied to the higher Total Number of Utterances.

Those mothers who had had three pregnancies or more at baseline were rated higher on Book Reading Quality than those with one or two pregnancies. This may be related to a pattern of older mothers being rated higher on Book Reading Quality.

Education and Literacy. Mothers who had completed 11th grade or above at baseline scored higher on Book Reading Quality. Mothers with higher TABE scores at baseline had higher Book Reading Quality ratings than those with the lowest TABE scores (7th grade or below). It is not surprising that mothers with more formal education and higher literacy skills are rated higher on this measure of reading skill and comfort with the task.

Psychosocial Characteristics. Mothers expressing a greater sense of personal control at baseline scored highest on Book Reading Quality. Mothers who feel a greater sense of control in their own lives may be more successful in acquiring literacy skills and more confident as educators of their young children.

4. **18-Month Follow-Up Characteristics and Book Reading Performance.** We turn next to the relationship between maternal and family characteristics at the time of the 18-month follow-up and book reading performance, seeking possible differences in book reading between groups of mothers differentiated by demographic, educational, and psychosocial characteristics at this later time (see Table 5.3).

Demographic Characteristics. Mothers' talk during book reading (Number of Utterances, Percentage of Immediate Utterances, and Number of Discussion Topics) is strongly associated with children's age. This corresponds to results of previous research on book reading (De Temple, 1994). There is better performance with older children. When reading with younger children mothers provide support for the text of the book with immediate talk such as paraphrasing, drawing their attention to illustrations, and checking their comprehension with factual questions. Additionally, mothers may need to use more talk related to behavior management to help the children remain engaged in the activity. The greater Number of

Table 5.3

Means on Maternal Book Reading Task Variables, by 18-Month Follow-Up Characteristics

Subgroup Formed at 18 Months	n	Total Number of Utterances	Number of Non-immediate Utterances	Percentage of Immediate Utterances	Number of Discussion Topics	Book Reading Quality
Full Sample						
Mean		50.76	1.94	68.3	2.56	6.54
Standard deviation		36.45	2.64	12.87	1.42	1.48
Range		1-195	0-17	18.50-100	0-7	3-9
Demographic Characteristics						
Child's age at observational study (months)						
24-36	86	59.77 ***	1.58	69.34 *	2.20 **	6.33
37-47	115	47.81	1.98	69.59	2.74	6.60
48-63	89	40.64	2.22	65.88	2.84	6.64
Residence						
Living with partner/husband	55	44.87	1.75	65.97	2.58	6.33
Not living with partner/husband	235	49.38	2.00	68.77	2.64	6.59
Living with parent/grandparent	101	48.97	1.84	70.33 *	2.54	6.45
Not living with parent/grandparent	189	47.66	2.02	67.17	2.67	6.59
Education and Literacy						
TABE reading score (grade equivalent)[a]						
7th grade or below	138	52.30	2.02	68.92	2.49	6.08 ***
8th or 9th	70	47.47	1.87	68.20	2.73	6.77
10th or above	80	44.50	1.99	67.24	2.70	7.01 ***
Mother's education at 18 months						
No high school diploma or GED, not in school	151	51.60	1.69	69.22	2.64	6.13 ***
In process of getting GED or high school diploma	29	40.44	2.15	65.75	2.60	6.64
Has GED	55	42.71	2.01	66.80	2.41	6.82
High school/trade school diploma or attending college	55	50.12	2.49	78.33	2.81	7.27

(continued)

Table 5.3 (continued)

Subgroup Formed at 18 Months	n	Total Number of Utterances	Number of Non-immediate Utterances	Percentage of Immediate Utterances	Number of Discussion Topics	Book Reading Quality
Earnings from Employment						
Respondent's total earnings over the 18-month period[b]						
No earnings	167	52.88 *	2.06	68.09	2.48	6.40
Below mean	86	42.06	1.63	67.77	2.88	6.64
At or above mean ($3,323.87)	37	42.97	2.18	69.88	2.75	6.95
Total earnings from respondent and partner or others during month 18[b]						
No earnings	212	52.33 *	2.03	68.10	2.62	6.45
Below mean	43	33.82	1.96	67.80	2.76	6.95
At or above mean ($719.40)	35	43.88	1.52	69.61	2.49	6.59
Fertility						
Pregnancy during follow-up	165	50.70	1.95	68.91	2.53	6.32 ***
No pregnancy during follow-up	125	45.95	1.96	67.44	2.74	6.81
Birth during follow-up	103	54.31 *	2.01	69.19	2.54	6.29 *
No birth during follow-up	187	45.67	1.93	67.78	2.67	6.67
Abortion during follow-up	42	49.99	1.69	70.30	2.58	6.22
No abortion during follow-up	248	48.25	2.01	67.85	2.64	6.60
Two or more pregnancies during follow-up	42	49.81	2.23	70.14	2.41	6.38
Did not have two or more pregnancies during follow-up	248	48.29	1.91	67.90	2.67	6.57
Sexually active at follow-up	24	47.74	1.93	67.78	2.62	6.56
Not sexually active at follow-up	44	51.08	1.93	70.46	2.69	6.49

(continued)

BEST COPY AVAILABLE

Table 5.3 (continued)

Subgroup Formed at 18 Months	n	Total Number of Utterances	Number of Non-immediate Utterances	Percentage of Immediate Utterances	Number of Discussion Topics	Book Reading Quality
Child Care[c]						
Any child care during follow-up	253	46.82 ***	1.97	66.19	2.70 **	6.60 *
No child care during follow-up	27	68.09	2.28	68.46	1.96	6.05
Child ever in child care before age 1	86	45.99	1.79	68.33	2.47	6.41
Child not in care before age 1	194	50.06	2.09	68.21	2.69	6.60
Number of months child was in day care/preschool						
Below mean	182	48.08	1.77	68.65	2.64	6.30
At or above mean (3.7 months)	108	49.25	2.26	67.58	2.60	6.93
Number of months child was in care by grandparent						
Below mean	203	50.06	1.99	67.12 **	2.55	6.58
At or above mean (2.9 months)	87	45.13	1.89	70.73	2.79	6.47
Psychosocial Characteristics						
CES-D (depression) score[d]						
0-15 (not at risk)	177	49.21	2.02	68.69 *	2.69	6.71 *
16-23 (at some risk)	67	46.58	1.88	70.03	2.69	6.35
24-60 (at high risk)	46	48.42	1.82	64.05	2.27	6.16
Mastery score[e]						
Below mean	117	47.42	1.64	66.35	2.39 *	6.39
At or above mean (22)	173	49.27	2.17	69.51	2.78	6.64
Difficult Life Circumstances score[f]						
Below mean	143	47.82	2.02	67.95	2.78 *	6.62
At or above mean (2.8)	147	49.20	1.89	68.52	2.47	6.47
Number of social supports[g]						
0-2	202	49.71	1.98	67.57	2.62	6.44 *
3 or more	87	45.28	1.85	69.76	2.64	6.79

(continued)

Table 5.3 (continued)

Subgroup Formed at 18 Months	n	Total Number of Utterances	Number of Non-immediate Utterances	Percentage of Immediate Utterances	Number of Discussion Topics	Book Reading Quality
Contact with Father of Child						
Father babysat for child in past year	127	50.74	2.02	68.52	2.57	6.58
Father did not babysit for child in past year	162	46.87	1.91	68.06	2.67	6.52
Father took child overnight	139	51.06	1.99	68.07	2.50	6.58
Father did not take child overnight	150	46.35	1.92	68.44	2.74	6.51
Father saw child once or not at all	82	39.85 **	1.65	67.87	2.58	6.29 *
Father saw child more than once	207	51.99	2.08	68.42	2.65	6.65
Substance Abuse						
Respondent ever high on alcohol in past month	106	45.00	1.81	67.57	2.70	6.37
Respondent never high on alcohol in past month	179	50.38	1.99	68.79	2.63	6.67
Respondent used drugs in past month[h]	43	50.39	1.70	71.25	2.52	6.06 *
Respondent did not use drugs in past month[h]	242	48.15	1.97	67.83	2.67	6.67
Sample size	290					

(continued)

Table 5.3 (continued)

Sources: New Chance 18-month follow-up survey and coded observational study variables.

Notes: Means are adjusted for experimental/control status in order to be descriptive of the overall sample.

Statistical significance levels are indicated as *** < = 1 percent, ** < = 5 percent, and * < = 10 percent, which refers to the probability that sample means are different from each other only because of random error.

For subgroups that have more than two levels, the placement of asterisks does not indicate where the significant difference lies, but only that there was a significant difference among the set of means (but see note a below).

Calculations for this table used data for all 290 respondents for whom there were 18 months of follow-up survey data, and observational data coded for literacy, including those with values of zero for outcomes and New Chance enrollees (i.e., experimentals) who did not participate in the program. The sample size may fall slightly short of the number reported because of missing or unusable items for some respondents' questionnaires and videotape problems.

[a] Scores on each grade-equivalent level shown are compared with scores in the group that was not at that level (e.g., 8th or 9th grade vs. not in 8th or 9th grade); thus, where the difference is statistically significant, asterisks are shown for each level of grade equivalence.

[b] Means were computed excluding values of zero.

[c] Regular child care was defined as an ongoing arrangement used while the mother was in school, in training, or working.

[d] The Center for Epidemiological Studies Depression (CES-D) Scale is a widely used measure of depression; scores can range from zero to 60.

[e] The Mastery Scale measures sense of mastery over personal events; scores can range from 7 to 28.

[f] Scores represent total numbers of ongoing problems or stresses the respondent faces, of a list of 10 problems.

[g] Respondents were asked to indicate the total number of individuals to whom they could turn for support.

[h] Self-report measures that included use of marijuana, cocaine, crack, heroin, PCP, and ice.

Discussion Topics used with older children reflects the ability of older children to sustain longer and more complex discussions.

Education and Literacy. Mothers with higher literacy scores at the time of the 18-month follow-up perform better on book reading measures. Predictably, the difference is significant in Book Reading Quality, a variable that includes mother's reading fluency. Educational attainment at the 18-month follow-up also predicts Book Reading Quality. The lowest book reading performance is by those without a diploma who are not currently in school and the highest performance is by those with a high school diploma or attending college.

Mothers who were not pregnant during the follow-up period had higher Book Reading Quality scores than those who were pregnant. Mothers who did not give birth during the follow-up had a lower Total Number of Utterances and higher Book Reading Quality scores than those who did give birth during the follow-up period.

Child Care. Mothers with some child care during the 18-month follow-up period talked less (a pattern associated with reading to older children), had a higher Number of Discussion Topics, and scored higher on Book Reading Quality than those without child care. Many aspects of child care may positively influence mothers' book reading with their children — for example, observation of child care providers reading or talking with children or direct suggestions about reading with children at home. Children may also influence their mothers' reading through skills they acquire in child care (for example, paying attention, asking questions).

Psychosocial Characteristics. Mothers at lower risk for depression tended to do better in book reading, with significantly higher scores on Book Reading Quality. Having a greater number of social supports was also associated with higher Book Reading Quality scores.

Contact with the fathers of the children was associated with mothers' book reading. Mothers had higher Book Reading Quality scores and higher Total Number of Utterances if fathers saw their children more than once. Total Number of Utterances is strongly associated with children's age. Younger children may have had more contact with fathers than older children. Better book reading performance is also accomplished by mothers who had not used drugs in the previous month; this difference was significant for the Book Reading Quality score.

5. Evidence for Program Impacts on Mother-Child Interaction During Book Reading and Discussion. We will now examine the impact of New Chance on mothers' book reading by comparing book reading characteristics of mothers assigned to the New Chance experimental group and those assigned to the control group.

Mothers in the experimental group were rated more highly than those in the control group on the global measure of Book Reading Quality, reflecting the composite of ratings on mother's Comfort Level, Reading Fluency, and Reading Intonation and Animation (see Table 5.4).

6. Program Impacts for Subgroups. We turn now to an investigation of program impacts on aspects of book reading for baseline subgroups (see Tables 5.5–5.9). In none of these subgroup comparisons did the control group have significantly more favorable scores than the experimental group on the book reading variables. In the case of the variable Percentage of

Table 5.4

Impacts of New Chance on Maternal Book Reading Task Variables

Observational Variable	Experimentals	Controls	Difference	p[a]
Total Number of Utterances	50.67	50.93	-0.26	0.954
Number of Nonimmediate Utterances	2.11	1.66	0.45	0.172
Percentage of Immediate Utterances	68.00	68.85	-0.85	0.584
Number of Discussion Topics	2.65	2.41	0.24	0.173
Book Reading Quality	6.67	6.32	0.35 *	0.064
Sample size	184	106		

Source: New Chance coded observational study variables.

Notes: Calculations for this table used data for all 290 respondents for whom there were observational data coded for literacy and 18-month follow-up survey data including those with values of zero for outcomes and New Chance enrollees (i.e., experimentals) who did not participate in the program. The sample sizes may fall slightly short of the numbers reported because of videotape or audiotape problems.

The averages or percentages are adjusted using linear analysis of covariance procedures controlling for seven kinds of difference in characteristics before random assignment (age of child, maternal literacy, whether mother had more than one child, gender of child, race/ethnicity, Philadelphia site, and Portland site).

[a] An F-test was applied to each difference in regression-adjusted means. The column labeled "p" is the statistical significance level of each between-group impact. That is, p is the probability that the averages for the experimental and control groups are different from each other only because of random error. Statistical significance levels are indicated as ***<=1 percent; **<=5 percent; *<= 10 percent.

Table 5.5

Impacts of New Chance on Mother's Total Number of Utterances During Book Reading, by Subgroups Formed at Baseline

Characteristics and Subgroup at Baseline	n	Total Number of Utterances		Within-Subgroup Impact	p[a]	Between-Subgroups Difference[b]	p[c]
		Experimentals	Controls				
Full Sample							
Mean		50.67	50.93	-0.26	0.954		
Demographic Characteristics							
Child's age at observational study (months)							
24-36	86	47.61	46.78	0.83	0.919	-----	0.936
37-47	115	48.40	50.36	-1.96	0.784		
48-63	89	56.94	55.22	1.72	0.829		
Gender of child[d]						2.51	0.775
Female	142	49.00	48.03	0.97	0.875		
Male	148	52.30	53.84	-1.54	0.808		
Number of children						-8.14	0.371
1	183	48.68	52.15	-3.47	0.544		
2 or more	107	54.39	49.72	4.67	0.509		
Race/ethnicity						4.15	0.771
Black	244	51.81	51.62	0.19	0.967		
White	46	44.84	48.80	-3.96	0.769		
Age of mother (years)						----- ***	0.008
16-17	56	66.15	43.94	22.21	0.025		
18-19	148	41.13	53.44	-12.31 **	0.044		
20-22	86	56.95	52.15	4.8 **	0.544		
Living arrangement (mother)						3.92	0.666
Living with mother	104	43.88	42.23	1.65	0.816		
Not living with mother	179	54.69	56.96	-2.27	0.689		
Living arrangement (partner/husband)						14.47	0.339
Living with partner/husband	30	60.73	47.55	13.18	0.370		
Not living with partner/husband	253	49.70	51.26	-1.56	0.741		
Age at first child's birth (years)						8.71	0.327
13-16	135	53.11	48.88	4.23	0.506		
17-19	155	48.69	53.17	-4.48	0.472		
Age of youngest child (years)						8.07	0.371
Under 1	107	45.12	40.83	4.29	0.541		
1 or older	183	53.86	57.64	-3.78	0.505		
Number of pregnancies						12.16	0.235
1 or 2	218	50.93	48.12	2.81	0.397		
3 or more	71	49.47	58.82	-9.35	0.582		

(continued)

Table 5.5 (continued)

Characteristics and Subgroup at Baseline	n	Total Number of Utterances Experimentals	Total Number of Utterances Controls	Within-Subgroup Impact	p[a]	Between-Subgroups Difference[b]	p[c]
Education and Literacy							
Highest grade completed						16.09 *	0.070
10th or below	185	51.79	45.73	6.06	0.279		
11th or above	105	48.68	59.01	-10.03	0.149		
Interval since last attended regular high school						-6.98	0.432
2 years or less	134	47.34	51.25	-3.91	0.545		
3 years or more	149	53.77	50.70	3.07	0.618		
TABE reading test score[e]						-----	0.933
7th grade or below	132	49.67	51.43	-1.76	0.791		
8th or 9th grade	79	54.10	54.93	-0.83	0.921		
10th grade or above	79	48.81	46.72	2.09	0.799		
Employment and AFDC History							
Ever employed						-0.42	0.969
Yes	244	50.13	50.42	-0.29	0.955		
No	66	52.61	52.48	0.13	0.989		
Family received AFDC when sample member was growing up						-----	0.270
Always	60	56.44	46.97	9.47	0.315		
Sometimes	145	49.56	56.21	-6.65	0.286		
Never	84	49.05	43.34	5.71	0.510		
Psychosocial Characteristics							
CES-D Scale (depression)[f] (%)						-----	0.754
0-15 (not at risk)	141	53.26	50.53	2.73	0.667		
16-23 (at some risk)	77	50.87	56.10	-5.23	0.542		
24-60 (at high risk)	72	45.50	45.89	-0.39	0.964		
Self-Esteem score[g]						11.24	0.265
Below the mean	75	43.25	51.93	-8.68	0.322		
At or above the mean (35)	215	53.29	50.73	2.56	0.616		
Number of sources of emotional support						14.50 *	0.099
0-2	133	53.93	46.72	7.21	0.255		
3 or more	157	48.24	55.53	-7.29	0.238		
Level of satisfaction with emotional support[h]						5.03	0.566
Less than very satisfied	138	51.85	49.46	2.39	0.708		
Very satisfied	152	49.69	52.33	-2.64	0.666		

(continued)

BEST COPY AVAILABLE

Table 5.5 (continued)

Characteristics and Subgroup at Baseline	n	Total Number of Utterances Experimentals	 Controls	Within-Subgroup Impact	p[a]	Between-Subgroups Difference[b]	p[c]
Locus of Control score[i]						2.57	0.755
Below the mean	115	47.85	49.74	-1.69	0.784		
At or above the mean (22)	175	52.55	51.67	0.88	0.877		
Sample size		184	106				

Sources: New Chance baseline enrollment data and coded observational study variables.

Notes: Calculations for this table used data for all 290 respondents for whom there were observational data coded for literacy and 18 months of follow-up survey data, including those with values of zero for outcomes and New Chance enrollees (i.e., experimentals) who did not participate in the program. The sample sizes may fall slightly short of this number because of missing or unusable items from some respondents' questionnaires or videotape problems.

The averages are adjusted using linear analysis of covariance procedures controlling for up to seven kinds of difference in characteristics before random assignment (age of child, maternal literacy, whether or not mother had more than one child, gender of child, race/ethnicity, Philadelphia site, and Portland site).

[a] A two-tailed t-test was applied to each difference in regression-adjusted means. The column labeled "p" is the statistical significance level of each within-subgroup impact. That is, p is the probability that sample estimates are different from zero or from each other only because of random error. Statistical significance levels are indicated as ***<=1 percent; **<=5 percent; *<= 10 percent.

[b] For each characteristic with only two subgroups, the between-subgroups difference is the impact for the first subgroup less the impact for the second subgroup. For characteristics with more than two subgroups, no single difference between subgroup impacts can be calculated, as indicated by dashes in the table. However, it is possible to assess the statistical significance of variation across multiple subgroups, as indicated by the asterisks.

[c] An F-test was applied to the interaction between the subgroup and experimental/control status. The column labeled "p" is the statistical significance level of each between-subgroups difference in impacts. That is, p is the probability that sample estimates are different from zero or from each other only because of random error. Statistical significance levels are indicated as ***<=1 percent; **<=5 percent; *<= 10 percent.

[d] When a sample member had more than one child, her response refers to the randomly selected focal child.

[e] The test used to measure reading ability was the reading part of the Tests of Adult Basic Education (TABE). Most sites administered the Survey Form of the test, but some administered the full reading test.

[f] The Center for Epidemiological Studies Depression (CES-D) Scale is a widely used measure of depression; scores can range from zero to 60.

[g] The measure of self-esteem used was the Rosenberg Self-Esteem Scale, a 10-item scale that assesses a person's global sense of self-worth. Scores can range from 10 to 50; 30 is considered the neutral midpoint.

[h] Enrollees were asked about their degree of satisfaction with the emotional support ("people who listen to you, reassure you, and show you they care") they received. Levels range from 1 (very dissatisfied) to 5 (very satisfied).

[i] The Locus of Control Scale is a six-item adaptation of the longer scale originally developed by Julien Rotter (1966). Scores can range from 6 to 30; 18 is considered the neutral midpoint.

Table 5.6

Impacts of New Chance on Mother's Number of Nonimmediate Utterances During Book Reading, by Subgroups Formed at Baseline

Characteristics and Subgroup at Baseline	n	Total Number Nonimmediate Utterances Experimentals	Total Number Nonimmediate Utterances Controls	Within-Subgroup Impact		p^a	Between-Subgroups Difference^b		p^c
Full Sample									
Mean		2.11	1.66	0.45		0.172			
Demographic Characteristics									
Child's age at observational study (months)							-----		0.928
24-36	86	1.76	1.39	0.37		0.536			
37-47	115	2.19	1.83	0.36		0.502			
48-63	89	2.34	1.71	0.63		0.281			
Gender of child^d							0.73		0.268
Female	142	2.17	1.36	0.81	*	0.080			
Male	148	2.06	1.98	0.08		0.860			
Number of children							-1.26	*	0.061
1	183	1.95	1.99	-0.04		0.909			
2 or more	107	2.44	1.22	1.22	**	0.021			
Race/ethnicity							-0.12		0.911
Black	244	2.14	1.70	0.44		0.211			
White	46	1.97	1.41	0.56		0.578			
Age of mother (years)							-----	*	0.058
16-17	56	2.74	1.39	1.35	*	0.069			
18-19	148	1.48	1.79	0.31		0.495			
20-22	86	2.80	1.67	1.13	*	0.057			
Living arrangement (mother)							-0.76		0.264
Living with mother	104	1.60	1.65	-0.05		0.924			
Not living with mother	179	2.41	1.70	0.71	*	0.096			
Living arrangement (partner/husband)							0.56		0.628
Living with partner/husband	30	2.00	1.05	0.95		0.387			
Not living with partner/husband	253	2.14	1.75	0.39		0.266			
Age at first child's birth (years)							-0.11		0.869
13-16	135	1.81	1.43	0.38		0.419			
17-19	155	2.36	1.87	0.49		0.288			
Age of youngest child (years)							1.41	**	0.035
Under 1	107	2.51	1.20	1.31	**	0.013			
1 or older	183	1.90	2.00	-0.10		0.797			
Number of pregnancies							0.31		0.673
1 or 2	218	2.25	1.71	0.54		0.152			
3 or more	71	1.68	1.45	0.23		0.735			
Education and Literacy									
Highest grade completed							0.38		0.579
10th or below	185	2.13	1.53	0.60		0.155			
11th or above	105	2.08	1.86	0.22		0.678			

(continued)

Table 5.6 (continued)

Characteristics and Subgroup at Baseline	n	Total Number Nonimmediate Utterances Experimentals	Controls	Within-Subgroup Impact	p[a]	Between-Subgroups Difference[b]	p[c]
Interval since last attended regular high school						-0.71	0.279
2 years or less	134	1.73	1.65	0.08	0.863		
3 years or more	149	2.46	1.67	0.79 *	0.081		
TABE reading test score[e]						-----	0.333
7th grade or below	132	1.74	1.77	-0.03	0.952		
8th or 9th grade	79	2.50	1.85	0.65	0.295		
10th grade or above	79	2.41	1.31	1.10 *	0.073		
Employment and AFDC History							
Ever employed						-0.06	0.932
Yes	224	2.05	1.60	0.45	0.241		
No	66	2.33	1.82	0.51	0.460		
Family received AFDC when sample member was growing up						-----	0.271
Always	60	1.97	1.91	0.06	0.933		
Sometimes	145	1.84	1.70	0.14	0.763		
Never	84	2.62	1.29	1.33 **	0.040		
Psychosocial Characteristics							
CES-D Scale (depression) (%)[f]						-----	0.888
0-15 (not at risk)	141	2.15	1.56	0.59	0.213		
16-23 (at some risk)	77	2.07	1.87	0.20	0.747		
24-60 (at high risk)	72	2.06	1.62	0.44	0.508		
Self-Esteem score[g]						0.68	0.368
Below the means	75	1.73	1.78	-0.05	0.932		
At or above the mean (35)	215	2.25	1.62	0.63	0.103		
Number of sources of emotional support						0.25	0.706
0-2	133	2.28	1.68	0.60	0.208		
3 or more	157	1.98	1.63	0.35	0.449		
Level of satisfaction with emotional support[h]						0.01	0.980
Less than very satisfied	138	2.05	1.60	0.45	0.341		
Very satisfied	152	2.16	1.72	0.44	0.337		
Locus of Control score[i]						-0.08	0.901
Below the mean)	115	1.94	1.44	0.50	0.331		
At or above the mean (22)	175	2.22	1.80	0.42	0.327		
Sample Size		184	106				

Sources and Notes: See Table 5.5.

Table 5.7

Impacts of New Chance on Mother's Percentage of Immediate Utterances During Book Reading, by Subgroups Formed at Baseline

Characteristics and Subgroup at Baseline	n	Percentage of Immediate Utterances: Experimentals	Percentage of Immediate Utterances: Controls	Within-Subgroup Impact	p[a]	Between-Subgroups Difference	p[b]
Full Sample							
Mean		68.00	68.85	-0.85	0.584		
Demographic Characteristics							
Child's age at observational study (months)						-----	0.184
24-36	86	66.18	68.81	-2.63	0.361		
37-47	115	69.99	67.54	2.45	0.328		
48-63	89	66.78	70.84	-4.06	0.148		
Gender of child[c]						1.73	0.575
Female	142	69.32	69.33	-0.01	0.998		
Male	148	66.74	68.48	-1.74	0.434		
Number of children						0.19	0.956
1	183	68.40	69.18	-0.78	0.696		
2 or more	107	67.29	68.26	-0.97	0.700		
Race/ethnicity						-0.51	0.919
Black	244	67.73	68.64	-0.91	0.583		
White	46	69.35	69.75	-0.40	0.933		
Age of mother (years)						-----	0.196
16-17	56	67.98	73.80	-5.82	0.103		
18-19	148	67.35	68.39	-1.04	0.636		
20-22	86	69.18	66.65	-2.53	0.376		
Living arrangement (mother)						-7.61 **	0.019
Living with mother	104	66.43	71.64	-5.21	0.040		
Not living with mother	179	68.96	66.56	2.40 **	0.235		
Living arrangement (partner/husband)						0.97	0.475
Living with partner/husband	30	66.34	67.51	-1.17	0.432		
Not living with partner/husband	253	68.63	68.83	-0.20	0.908		
Age at first child's birth (years)						-5.73 *	0.068
13-16	135	67.28	71.05	-3.77	0.093		
17-19	155	68.54	66.58	1.96 *	0.370		
Age of youngest child (years)						-4.16	0.195
Under 1	107	68.13	71.45	-3.32	0.184		
1 or older	183	67.87	67.03	0.84	0.679		
Number of pregnancies						0.91	0.781
1 or 2	218	68.35	68.85	-0.60	0.782		
3 or more	71	66.97	68.48	-1.51	0.633		
Education and Literacy							
Highest grade completed						-4.26	0.183
10th or below	185	67.93	70.46	-2.53	0.203		
11th or above	105	68.07	66.34	1.73	0.493		

(continued)

Table 5.7 (continued)

Characteristics and Subgroup at Baseline	n	Percentage of Immediate Utterances		Within-Subgroup Impact	p[a]	Between-Subgroups Difference	p[b]
		Experimentals	Controls				
Interval since last attended regular high school						0.47	0.882
2 years or less	134	68.19	69.15	-0.96	0.674		
3 years or more	149	68.28	68.77	-0.49	0.820		
TABE reading test score[d]						-----	0.147
7th grade or below	132	66.86	71.13	-4.27 *	0.069		
8th or 9th grade	79	68.95	66.81	2.14	0.470		
10th grade or above	79	69.11	67.61	1.50	0.604		
Employment and AFDC History							
Ever employed							
Yes	224	67.22	68.27	-1.50	0.555	-1.22	0.744
No	66	70.76	70.59	0.17	0.961		
Family received AFDC when sample member was growing up						-----	0.908
Always ·	60	68.29	68.16	0.13	0.968		
Sometimes	145	67.46	69.08	-1.62	0.466		
Never	84	68.64	69.56	-0.92	0.767		
Psychosocial Characteristics							
CES-D Scale (depression) (%)[e]						-----	0.608
0-15 (not at risk)	141	69.43	69.09	0.34	0.882		
16-23 (at some risk)	77	66.86	70.02	-3.16	0.270		
24-60 (at high risk)	72	66.43	66.77	-0.34	0.913		
Self-Esteem score[f]						3.49	0.324
Below the mean	75	63.84	67.31	-3.47	0.259		
At or above the mean (35)	215	69.45	69.43	0.02	0.993		
Number of sources of emotional support						-2.92	0.350
0-2	133	67.39	69.74	-2.35	0.296		
3 or more	157	68.44	67.87	0.57	0.797		
Level of satisfaction with emotional support[g]						3.17	0.305
Less than very satisfied	138	67.80	67.09	0.71	0.751		
Very satisfied	152	68.15	70.61	-2.46	0.254		
Locus of Control score[h]						8.60 ***	0.009
Below the mean	115	64.33	70.45	-6.12 **	0.018		
At or above the mean (22)	175	70.45	67.97	2.48	02.14		
Sample size		184	106				

(continued)

Table 5.7 (continued)

Sources: New Chance baseline enrollment data and coded observational study variables.

Notes: Calculations for this table used data for all 290 respondents for whom there were observational data coded for literacy and 18 months of follow-up survey data, including those with values of zero for outcomes and New Chance enrollees (i.e., experimentals) who did not participate in the program. The sample sizes may fall slightly short of this number because of missing or unusable items from some respondents' questionnaires or videotape problems.

The averages are adjusted using linear analysis of covariance procedures controlling for up to seven kinds of difference in characteristics before random assignment (age of child, maternal literacy, whether or not mother had more than one child, gender of child, race/ethnicity, Philadelphia site, and Portland site).

[a] A two-tailed t-test was applied to each difference in regression-adjusted means. The column labeled "p" is the statistical significance level of each within-subgroup impact. That is, p is the probability that sample estimates are different from zero or from each other only because of random error. Statistical significance levels are indicated as ***<=1 percent; **<=5 percent; *<= 10 percent.

[b] An F-test was applied to the interaction between the subgroup and experimental/control status. The column labeled "p" is the statistical significance level of each between-subgroups difference in impacts. That is, p is the probability that sample estimates are different from zero or from each other only because of random error. Statistical significance levels are indicated as ***<=1 percent; **<=5 percent; *<= 10 percent.

[c] When a sample member had more than one child, her response refers to the randomly selected focal child.

[d] The test used to measure reading ability was the reading part of the Tests of Adult Basic Education (TABE). Most sites administered the Survey Form of the test, but some administered the full reading test.

[e] The Center for Epidemiological Studies Depression (CES-D) Scale is a widely used measure of depression; scores can range from zero to 60.

[f] The measure of self-esteem used was the Rosenberg Self-Esteem Scale, a 10-item scale that assesses a person's global sense of self-worth. Scores can range from 10 to 50; 30 is considered the neutral midpoint.

[g] Enrollees were asked about their degree of satisfaction with the emotional support ("people who listen to you, reassure you, and show you they care") they received. Levels range from 1 (very dissatisfied) to 5 (very satisfied).

[h] The Locus of Control Scale is a six-item adaptation of the longer scale originally developed by Julien Rotter (1966). Scores can range from 6 to 30; 18 is considered the neutral midpoint.

Table 5.8

Impacts of New Chance on Number of Discussion Topics During Book Reading, by Subgroups Formed at Baseline

Characteristics and Subgroup at Baseline	n	Number of Discussion Topics		Within-Subgroup Impact		p[a]	Between-Subgroups Difference[b]		p[c]
		Experimentals	Controls						
Full Sample									
Mean		2.65	2.41	0.24		0.173			
Demographic Characteristics									
Child's age at observational study (months)							-----	*	0.093
24-36	86	2.40	2.53	-0.13		0.681			
37-47	115	3.00	2.31	0.69	**	0.012			
48-63	89	2.40	2.41	-0.01		0.974			
Gender of child[d]							0.56		0.102
Female	142	2.71	2.19	0.52	**	0.034			
Male	148	2.60	2.64	-0.04		0.843			
Number of children							0.13		0.700
1	183	2.66	2.37	0.29		0.193			
2 or more	107	2.62	2.46	0.16		0.578			
Race/ethnicity							-0.05		0.944
Black	244	2.64	2.41	0.23		0.206			
White	46	2.68	2.40	0.28		0.606			
Age of mother (years)							-----		0.843
16-17	56	2.60	2.57	0.03		0.938			
18-19	148	2.64	2.37	0.27		0.265			
20-22	86	2.69	2.38	0.31		0.330			
Living arrangement (mother)							0.42		0.242
Living with mother	104	2.69	2.21	0.48	*	0.088			
Not living with mother	179	2.60	2.54	0.06		0.787			
Living arrangement (partner/husband)							0.11		0.872
Living with partner/husband	30	3.08	2.76	0.32		0.587			
Not living with partner/husband	253	2.58	2.37	0.21		0.244			
Age at first child's birth (years)							-0.23		0.497
13-16	135	2.68	2.56	0.12		0.628			
17-19	155	2.61	2.26	0.35		0.143			
Age of youngest child (years)							-0.33		0.351
Under 1	107	2.25	2.23	0.02		0.928			
1 or older	183	2.86	2.51	0.35		0.111			
Number of pregnancies							0.28		0.491
1 or 2	218	2.65	2.36	0.29		0.147			
3 or more	71	2.63	2.62	0.01		0.965			
Education and Literacy									
Highest grade completed							-0.22		0.527
10th or below	185	2.60	2.44	0.16		0.485			
11th or above	105	2.73	2.35	0.38		0.178			

(continued)

Table 5.8 (continued)

Characteristics and Subgroup at Baseline	n	Number of Discussion Topics		Within-Subgroup Impact	p[a]	Between-Subgroups Difference[b]	p[c]
		Experimentals	Controls				
Interval since last attended regular high school						-0.20	0.560
2 years or less	134	2.81	2.69	0.12	0.650		
3 years or more	149	2.53	2.21	0.32	0.187		
TABE reading test score[e]						-----	0.799
7th grade or below	132	2.58	2.41	0.17	0.529		
8th or 9th grade	79	2.78	2.34	0.44	0.191		
10th grade or above	79	2.64	2.45	0.19	0.555		
Employment and AFDC History							
Ever employed						-0.32	0.425
Yes	224	2.67	2.51	0.16	0.433		
No	66	2.57	2.09	0.48	0.181		
Family received AFDC when sample member was growing up						-----	0.406
Always	60	2.61	2.57	0.04	0.907		
Sometimes	145	2.61	2.51	0.10	0.667		
Never	84	2.72	2.10	0.62 *	0.071		
Psychosocial Characteristics							
CES-D Scale (depression) (%)[f]						-----	0.676
0-15 (not at risk)	141	2.56	2.30	0.26	0.286		
16-23 (at some risk)	77	3.03	2.61	0.42	0.218		
24-60 (at high risk)	72	2.40	2.41	-0.01	0.990		
Self-Esteem score[g]						0.02	0.963
Below the mean	75	2.75	2.53	0.22	0.516		
At or above the mean (35)	215	2.61	2.37	0.24	0.228		
Number of sources of emotional support						0.07	0.851
0-2	133	2.67	2.39	0.28	0.276		
3 or more	157	2.63	2.42	0.21	0.394		
Level of satisfaction with emotional support[h]						0.23	0.507
Less than very satisfied	138	2.66	2.31	0.35	0.159		
Very satisfied	152	2.63	2.51	0.12	0.599		
Locus of Control score[i]						-0.24	0.541
Below the mean	175	2.56	2.42	0.14	0.521		
At or above the mean (22)	115	2.77	2.39	0.38	0.170		
Sample size		184	106				

Source and Notes: See Table 5.5

BEST COPY AVAILABLE

Table 5.9

Impacts of New Chance on Book Reading Quality, by Subgroups Formed at Baseline

Characteristics and Subgroup at Baseline	n	Book Reading Quality Experimentals	Book Reading Quality Controls	Within-Subgroup Impact	p[a]	Between-Subgroups Difference	p[b]
Full Sample		6.67	6.32	0.35 *	0.064		
Mean							
Demographic Characteristics							
Child's age at observational study (months)						-----	0.142
24-36	68	6.38	6.55	-0.17	0.632		
37-47	98	6.85	6.11	0.74 **	0.014		
48-63	88	6.66	6.42	0.24	0.443		
Gender of child[c]						0.63 *	0.081
Female	124	6.88	6.23	0.65	0.011		
Male	130	6.48	6.46	0.02 **	0.953		
Number of children						-0.10	0.779
One	155	6.61	6.31	0.30	0.222		
More than one	99	6.76	6.36	0.40	0.157		
Race/ethnicity						0.18	0.753
Black	216	6.67	6.30	0.37	0.065		
White	38	6.69	6.50	0.19 *	0.734		
Age of mother (years)						-----	0.377
16-17	44	6.74	6.01	0.73	0.109		
18-19	134	6.59	6.48	0.11	0.658		
20-22	76	6.78	6.24	0.54	0.114		
Living arrangement (mother)						-0.06	0.885
Living with mother	86	6.57	6.32	0.25	0.413		
Not living with mother	161	6.69	6.38	0.31	0.195		
Living arrangement (partner/husband)						-0.11	0.869
Living with partner/husband	25	6.51	6.31	0.20	0.745		
Not living with partner/husband	222	6.67	6.36	0.31	0.120		
Age at first child's birth (years)						0.00	0.993
13-16	123	6.74	6.39	0.35	0.189		
17-19	131	6.61	6.26	0.35	0.192		
Age of youngest child (years)						0.71 *	0.062
Under 1	93	6.80	6.02	0.78 **	0.009		
1 or older	161	6.61	6.54	0.07	0.778		
Number of pregnancies						0.09	0.830
1 or 2	190	6.61	6.24	0.37	0.083		
3 or more	63	6.85	6.57	0.28	0.452		
Education and Literacy							
Highest grade completed						-0.11	0.770
10th or below	160	6.59	6.28	0.31	0.192		
11th or above	94	6.81	6.39	0.42	0.162		

(continued)

Table 5.9 (continued)

Characteristics and Subgroup at Baseline	n	Book Reading Quality		Within-Subgroup Impact	p[a]	Between-Subgroups Difference	p[b]
		Experimentals	Controls				
Interval since last attended regular high school						0.42	0.251
2 years or less	117	6.92	6.38	0.54 **	0.042		
3 years or more	133	6.42	6.30	0.12	0.627		
TABE reading test score[d]						-----	0.613
7th grade or below	113	6.20	5.97	0.23	0.428		
8th or 9th grade	67	6.96	6.46	0.50	0.166		
10th grade or above	74	7.24	6.60	0.64 *	0.058		
Employment and AFDC History							
Ever employed						-0.13	0.764
Yes	199	6.70	6.39	0.31	0.146		
No	55	6.55	6.11	0.44	0.263		
Family received AFDC when sample member was growing up						-----	0.910
Always	56	6.79	6.48	0.31	0.425		
Sometimes	124	6.51	6.22	0.29	0.270		
Never	74	6.85	6.38	0.47	0.178		
Psychosocial Characteristics							
CES-D Scale (depression) (%)[e]						-----	0.293
0-15 (not at risk)	124	6.74	6.38	0.36	0.168		
16-23 (at some risk)	69	6.83	6.10	0.73 **	0.043		
24-60 (at high risk)	61	6.38	6.46	-0.08	0.825		
Self-Esteem score[f]						0.83 *	0.052
Below the mean	63	6.34	6.63	-0.29	0.437		
At or above the mean (35)	191	6.79	6.25	0.54 ***	0.010		
Number of sources of emotional support						0.16	0.660
0-2	113	6.78	6.34	0.44	0.102		
3 or more	141	6.60	6.32	0.28	0.272		
Level of satisfaction with emotional support[g]						0.51	0.160
Less than very satisfied	120	6.91	6.28	0.63 **	0.019		
Very satisfied	134	6.48	6.36	0.12	0.640		
Locus of Control score[h]						0.07	0.847
Below the mean	97	6.46	6.16	0.30	0.309		
At or above the mean (22)	157	6.80	6.43	0.37	0.116		
Sample Size		184	106				

Sources and Notes: See Table 5.7.

Immediate Utterances, a more favorable score is indicated by a lower percentage of this type of talk. Immediate Talk is less cognitively challenging and complex than Non-immediate Talk. It is desirable for mothers to use a relatively smaller proportion of Immediate Talk during book reading.

Total Number of Utterances, an indication of maternal contribution to the book reading interaction, showed an inconsistent pattern of experimental-control differences within maternal age groups; the 16- or 17-year-old mothers in the experimental group produced a higher Total Number of Utterances, whereas the 18- or 19-year-old mothers produced a lower total (see Table 5.5). No other subgroup effects were found.

Positive subgroup impacts on Number of Nonimmediate Utterances did occur for mothers reading to girls (but not to boys), for those with more than one child, for those who had had a child under age 1 at baseline, and for those not living with their mothers (see Table 5.6). New Chance also had a positive effect on Number of Nonimmediate Utterances for mothers who had been out of school longer, for those with the highest scores in literacy (TABE), and for those from families that had never received welfare. The youngest and oldest mothers assigned to the New Chance experimental group had a higher Number of Nonimmediate Utterances than their control group agemates, but no impact was seen for the mothers aged 18 and 19 at baseline. While somewhat scattered, all of these effects support the view that the New Chance Intervention had positive implications for mothers' literacy-related interactive behaviors, and several suggest that it may have been somewhat greater for slightly more advantaged participants — those with higher literacy scores, less of a welfare history, and the resources to live alone. The findings of positive impacts for mothers with more children and with children younger than the target child, though, are hard to reconcile with this interpretation.

Positive subgroup impacts on the Percentage of Immediate Utterances (that is, a decrease in proportion of immediate talk) during book reading occurred for mothers living with their own mothers at baseline, for those who were younger at the birth of their first child, for those with the lowest reading scores at baseline, and for those with lower than average sense of personal control (see Table 5.7). All these effects, thus, were found for relatively disadvantaged subgroups. This finding may seem to conflict with the findings for Number of Nonimmediate Utterances reported in Table 5.6, of positive impacts occurring particularly among less disadvantaged subgroups of mothers. However, if we see high-quality book reading interactions as having a number of components, some of which are easier to achieve than others, this paradox can be resolved; the simpler adjustment involves producing fewer Immediate Utterances (and thus more feedback and responses to the child), whereas producing more Nonimmediate Utterances requires more cognitive and social resources.

Positive subgroup impacts on the Number of Discussion Topics were observed for mothers reading with daughters, mothers living with their own mothers, and mothers whose families had never received welfare when they were growing up (see Table 5.8). It should be noted that discussion topics could be introduced by child or mother; thus, girls and children who have frequent contact with grandmothers may be generating higher scores by participating more effectively in the discussion themselves. There were no program impacts for subgroups based on maternal education or literacy.

Higher Book Reading Quality scores were assigned to experimental group mothers than to control group mothers in particular subgroups. Program impacts on Book Reading Quality scores occurred for mothers reading to daughters, black mothers, and those with a child under age 1 at baseline. Positive subgroup impacts on Book Reading Quality scores were seen for mothers who had been out of high school for two years or less (by contrast with those who had been out of high school for three years or more), for those with the highest reading scores at baseline, for those at some risk for depression, for those with self-esteem scores at or above the mean, and for those who were less than very satisfied with their emotional support. These findings are extremely scattered, with some more disadvantaged subgroups (those at risk of depression, those unsatisfied with emotional support, those with younger children) showing impacts, but on other variables less disadvantaged groups (those with higher reading scores, higher self-esteem) showing the impacts.

7. A Summary of Results on Book Reading and Discussion. It should be noted that the children were generally eager to be read to and interested in *The Very Hungry Caterpillar*. In a global rating of child involvement carried out by coders, only 4.5 percent of the children were rated low and the majority (63.7 percent) were rated highly involved.

Mothers displayed a wide range of book reading styles, as reflected in the amount and type of talk they used. The Total Number of Utterances, the Percentage of Immediate Utterances, and the Number of Discussion Topics were related to the age of the child. Mothers with higher educational attainment and higher literacy scores provided better book reading, as measured by Book Reading Quality scores, than those with lower levels of education and literacy. Better book reading performance was also associated with the use of child care and with scores indicating positive maternal psychosocial characteristics at 18 months.

Assignment to the New Chance experimental group had a positive impact on the overall quality of mothers' book reading as measured by the composite measure reflecting Reading Fluency, Reading Intonation and Animation, and Comfort Level with reading aloud to their children. The global measure Book Reading Quality reflects characteristics of book reading at a more basic level than the utterance level variables. It measures mothers' ability to read the text (fluency), their comfort with reading, and their comprehension and ability to interpret the text as demonstrated by intonation. Variations in the mothers' literacy levels and educational levels obviously relate to Book Reading Quality. Variables measuring the linguistic complexity of the interaction (such as Number of Nonimmediate Utterances) constitute a more fine-grained analysis, going *beyond* skill in carrying out the mechanics of reading aloud. The effectiveness of an intervention program aimed at mothers with very low levels of education and literacy may be best evaluated using the global rating rather than more intensive utterance level coding.

It may seem paradoxical that we found a program impact on Book Reading Quality when no impact was found on performance on the TABE, a test of literacy skill. It may be that the TABE is insensitive to differences in comfort level or oral reading skill: both are important in helping mothers read to their children. In the task of reading a fairly simple picture book to their children, mothers assigned to the New Chance experimental group performed better. In other words, while their word recognition and reading comprehension skills might not have been improved, their oral fluency evidently was. Standardized tests like the TABE do not reflect performance differences in the skill of reading aloud. It is striking that there are such large individual differences in performance in reading even the very simple text of *The Very Hungry Caterpillar.* Mothers who read more fluently and with better intonation (precisely the skills standardized tests do not assess) are much more likely to engage their children successfully in book reading activities. Substantial evidence confirms that parent-child book reading is one mechanism for the intergenerational transfer of literacy.

The mother's educational achievement could affect the quality of her reading, not only by improving her skills in the mechanics of reading, but also by providing her with experiences and strategies that would help her improve her own and thereby her child's skills in acquiring information, in comprehension, and in drawing inferences through reading. In order to have a greater impact on the mother and child's home literacy experience an intervention program must focus on the mother's education and the mother's knowledge of children's learning and development.

Parenting programs could address the importance of book reading with young children and thereby, at a minimum, affect the frequency with which book reading occurs. These programs would have to be sufficiently intensive and extensive to move beyond crisis management and behavioral issues to include means of enhancing cognitive stimulation.

B. The Wheels Task

1. Descriptive Analyses. We turn now to parallel results from the wheels task analyzing the three variables: Objects Named, number of wheeled objects named correctly by the child; Objects/Elicitations, wheeled objects named as a proportion of maternal elicitations; and Mother's Ease of Ideas. Again, we will progress from descriptive analyses of the three wheels task variables to impact analyses.

Objects Named ranged from none to 14, with a mean of 2.67 (see Table 5.10). Mothers typically had to ask their children several questions or provide several prompts for each correct response. Objects/Elicitations ranged from zero to 1.25, with a mean of 0.27. Mother's Ease of Ideas ranged from 1 to 4, with a mean of 2.50.

2. Baseline Characteristics and Wheels Task Performance: Demographic Characteristics. Wheels task performance was related to the age of the youngest child in the home at baseline. When the youngest child had been age 1 or older at baseline, the mothers and children did better on all three measures. It may be that the demands of a younger child diminish the opportunities for extended talk with the older child and thus hinder both the mother and child in carrying out this task. This may also be a function of the age of the target child. Older children and their mothers did better on all measures (see Table 5.11).

Education and Literacy. Mothers who had completed 11th grade or above at baseline were rated higher on Ease of Ideas and were more successful on the Objects/Elicitation measure than those who had completed only 10th grade or below (see Table 5.10). Mothers who were more successful in school and/or more familiar with school-like tasks (for example, giving formal definitions of words) would predictably be more comfortable and successful in helping children with this guessing game.

Table 5.10

Means on Wheels Task Variables, by Subgroups Formed at Baseline

Subgroup Formed at Baseline	n	Objects Named	Objects/ Elicitations	Mother's Ease of Ideas
Full Sample				
Mean		2.67	0.27	2.50
Standard deviation		2.61	0.26	0.92
Range		0-14	0-1.25	1-4
Demographic Characteristics				
Gender of child[a]				
Female	142	2.40 *	0.25	2.44
Male	148	2.93	0.29	2.56
Number of children				
1	1.83	2.75	0.27	2.50
2 or more	1.07	2.54	0.27	2.51
Race/ethnicity				
Black	244	2.57	0.27	2.48
White	46	3.21	0.29	2.65
Age of mother (years)				
16-17	56	2.36	0.24	2.39
18-19	148	2.60	0.27	2.46
20-22	86	3.00	0.29	2.66
Living arrangement (mother)				
Living with mother	104	2.35	0.25	2.40
Not living with mother	179	2.85	0.28	2.58
Living arrangement (partner/husband)				
Living with partner/husband	30	3.11	0.32	2.72
Not living with partner/husband	253	2.61	0.26	2.49
Age at first child's birth (years)				
13-16	135	2.91	0.28	2.51
17-19	155	2.46	0.26	2.50
Age of youngest child (years)				
Under 1	107	1.82 ***	0.20 ***	2.26 ***
1 or older	183	3.17	0.31	2.64
Number of pregnancies				
1 or 2	218	2.75	0.28	2.52
2 or more	71	2.46	0.23	2.45
Education and Literacy				
Highest grade completed				
10th or below	185	2.53	0.25 *	2.44 *
11th or above	105	2.93	0.31	2.62

(continued)

Table 5.10 (continued)

Subgroup Formed at Baseline	n	Objects Named	Objects/ Elicitations	Mother's Ease of Ideas
Interval since last attended regular high school				
2 years or less	134	2.76	0.28	2.51
3 years or more	149	2.58	0.26	2.50
TABE reading test score (grade equivalent)[b]				
7th grade or below	132	2.50	0.28	2.41
8th or 9th grade	79	2.75	0.27	2.53
10th grade or above	79	2.87	0.26	2.63
Employment and AFDC History				
Ever employed				
Yes	224	2.60	0.26	2.44 **
No	66	2.92	0.30	2.71
Family received AFDC when sample member was growing up				
Never	84	2.91	0.31 **	2.58
Sometimes	145	2.48	0.23	2.47
Always	60	2.81	0.32	2.47
Psychosocial Characteristics				
CES-D (depression) score[c]				
0-15 (not at risk)	141	2.62	0.27	2.53
16-23 (at some risk)	77	2.95	0.31	2.56
24-60 (at high risk)	72	2.49	0.24	2.39
Self-Esteem score[d]				
Below mean	75	2.75	0.27	2.44
At or above the mean (35)	215	2.65	0.27	2.52
Number of sources of emotional support				
0-2	133	2.59	0.26	2.44
3 or more	157	2.74	0.28	2.56
Level of satisfaction with emotional support[e]				
Less than very satisfied	138	2.71	0.28	2.51
Very satisfied	152	2.64	0.26	2.50
Locus of Control score[f]				
Below mean	115	2.43	0.25	2.36 **
At or above mean (22)	175	2.83	0.28	2.60
Sample size	290			

(continued)

Table 5.10 (continued)

Sources: New Chance baseline enrollment data and coded observational study variables.

Notes: Means are adjusted for experimental/control status in order to be descriptive of the overall sample. Statistical significance levels are indicated as *** < = 1 percent, ** < = 5 percent, and * < = 10 percent, which refers to the probability that sample means are different from each other only because of random error.

For subgroups that have more than two levels, the placement of asterisks does not indicate where the significant difference lies, but only that there was a significant difference among the set of means.

Calculations for this table used data for all 290 respondents for whom there were observational data and 18-month follow-up survey data, including those with values of zero for outcomes and New Chance enrollees (i.e., experimentals) who did not participate in the program. The sample size may fall slightly short of the number reported because of missing or unusable items from some respondents' questionnaires and videotape problems.

[a] When a sample member had more than one child, her response refers to the randomly selected focal child.

[b] The test used to measure reading ability was the reading part of the Tests of Adult Basic Education (TABE). Most sites administered the survey form of the test, but some administered the full reading test.

[c] The Center for Epidemiological Studies Depression (CES-D) scale is a widely used measure of depression; scores can range from zero to 60.

[d] The measure of self-esteem used was the Rosenberg Self-Esteem Scale, a 10-item scale that assesses a person's global sense of self-worth. Scores can range from 10 to 50; 30 is considered the neutral midpoint.

[e] Enrollees were asked about their degree of satisfaction with the emotional support ("people who listen to you, reassure you, and show you they care") they received. Levels range from 1 (very dissatisfied) to 5 (very satisfied); however, because so few individuals indicated that they were less than very satisfied with their emotional support, levels 1 through 4 were collapsed into one category labeled "less than very satisfied."

[f] The Locus of Control Scale is a 6-item adaptation of the longer scale originally developed by Julien Rotter (1966). Scores can range from 6 to 30; 18 is considered the neutral midpoint.

BEST COPY AVAILABLE

Table 5.11

Means on Wheels Task, by 18-Month Follow-Up Characteristics

Subgroup Formed at 18 Months	n	Objects Named	Objects/ Elicitations	Mother's Ease of Ideas
Full Sample				
Mean		2.67	0.27	2.5
Standard deviation		2.61	0.26	0.92
Range		0-14	0-1.25	1-4
Demographic Characteristics				
Child's age at observational study (months)				
24-36	86	1.13 ***	0.16 ***	2.20 ***
37-47	115	2.42	0.25	2.47
48-63	89	4.49	0.41	2.84
Residence				
Living with partner/husband	55	2.67	0.29	2.56
Not living with partner/husband	235	2.67	0.27	2.49
Living with parent/grandparent	101	2.98	0.28	2.49
Not living with parent/grandparent	189	2.51	0.26	2.51
Education and Literacy				
TABE reading score (grade equivalent)[a]				
7th grade or below	138	2.36 *	0.26	2.34 ***
8th or 9th	70	3.03	0.32 *	2.62
10th or above	80	2.84	0.24	2.65 *
Mother's education at 18 months				
No high school diploma or GED, not in school	151	2.44	0.26	2.37 **
In process of getting GED or high school diploma	29	2.93	0.35	2.75
Has GED	55	2.58	0.23	2.51
High school or trade school diploma/attending college	55	3.27	0.29	2.73
Earnings from Employment				
Respondent's total earnings over the 18-month period[b]				
No earnings	37	2.41 *	0.29	2.49
Below mean	86	3.23	0.30	2.67
At or above mean ($3,323.87)	167	2.57	0.25	2.42
Total earnings from respondent and partner or others during month 18[b]				
No earnings	35	2.66	0.27	2.51
Below mean	43	3.37	0.31	2.63
At or above mean ($719.40)	212	2.53	0.26	2.48

(continued)

Table 5.11 (continued)

Subgroup Formed at 18 Months	n	Objects Named	Objects/ Elicitations	Mother's Ease of Ideas
Fertility				
Pregnancy during follow-up	165	2.65	0.28	2.50
No pregnancy during follow-up	125	2.71	0.26	2.51
Birth during follow-up	103	2.56	0.28	2.45
No birth during follow-up	187	2.74	0.26	2.53
Abortion during follow-up	42	2.47	0.24	2.45
No abortion during follow-up	248	2.71	0.28	2.51
Two or more pregnancies during follow-up	42	2.50	0.25	2.38
Did not have two or more pregnancies during follow-up	248	2.70	0.27	2.52
Sexually active at follow-up	244	2.67	0.27	2.51
Not sexually active at follow-up	44	2.76	0.27	2.55
Child Care[c]				
Any child care during follow-up	253	2.76 *	0.28 *	2.52
No child care during follow-up	27	1.77	0.18	2.29
Child ever in child care before age 1	86	2.69	0.26	2.52
Child not in care before age 1	194	2.65	0.28	2.49
Number of months child was in day care/preschool				
Below mean	182	2.45 *	0.25	2.46
At or above mean (3.7 months)	108	3.05	0.30	2.58
Number of months child was in care by grandparent				
Below mean	203	2.55	0.26	2.47
At or above mean (2.9 months)	87	2.96	0.30	2.58
Psychosocial Characteristics				
CES-D (depression) score[d]				
0-15 (not at risk)	177	2.64	0.26	2.54
16-23 (at some risk)	67	3.02	0.30	2.49
24-60 (at high risk)	46	2.27	0.25	2.39
Mastery score[e]				
Below mean	117	2.26 **	0.25	2.39 *
At or above mean (22)	173	2.95	0.28	2.58
Difficult Life Circumstances score[f]				
Below mean	143	2.78	0.27	2.50
At or above mean (2.8)	147	2.57	0.27	2.50
Number of social supports[g]				
0-2	202	2.59	0.27	2.47
3 or more	87	2.88	0.28	2.60

(continued)

Table 5.11 (continued)

Subgroup Formed at 18 Months	n	Objects Named	Objects/ Elicitations	Mother's Ease of Ideas
Contact with Father of Child				
Father babysat for child in past year	127	2.50	0.26	2.42
Father did not babysit for child in past year	162	2.83	0.28	2.57
Father took child overnight	139	2.42	0.25 *	2.37 **
Father did not take child overnight	150	2.92	0.30	2.63
Father saw child once or not at all	82	2.89	0.31 *	2.64
Father saw child more than once	207	2.60	0.26	2.45
Substance Abuse				
Respondent ever high on alcohol in past month	106	2.50	0.25	2.52
Respondent never high on alcohol in past month	179	2.76	0.28	2.51
Respondent used drugs in past month[h]	43	2.58	0.27	2.46
Respondent did not use drugs in past month	242	2.72	0.27	2.53
Sample size	290			

Sources: New Chance 18-month follow-up survey and coded observational study variables.

Notes: Means are adjusted for experimental/control status in order to be descriptive of the overall sample.

Statistical significance levels are indicated as *** < = 1 percent, ** < = 5 percent, and * < = 10 percent, which refers to the probability that sample means are different from each other only because of random error.

For subgroups that have more than two levels, the placement of asterisks does not indicate where the significant difference lies, but only that there was a significant difference among the set of means.

Calculations for this table used data for all 290 respondents for whom there were observational data and 18-month follow-up survey data, including those with values of zero for outcomes and New Chance enrollees (i.e., experimentals) who did not participate in the program. The sample size may fall slightly short of the number reported because of missing or unusable items for some respondents' questionnaires and videotape problems.

[a] Scores on each grade equivalent level shown are compared with scores in the group that was not at that level (e.g., 8th or 9th grade vs. not in 8th or 9th grade); thus, where the difference is statistically significant, asterisks are shown for each level of grade equivalence.

[b] Means were computed excluding values of zero.

[c] Regular child care was defined as an ongoing arrangement used while the mother was in school, in training, or working.

[d] The Center for Epidemiological Studies Depression (CES-D) Scale is a widely used measure of depression; scores can range from zero to 60.

[e] The Mastery Scale measures sense of mastery over personal events. Scores can range from 7 to 28.

[f] Scores represent total numbers of ongoing problems or stresses the respondent faces, of a list of 10 problems.

[g] Respondents were asked to indicate the total number of individuals to whom they could turn for support.

[h] Self-report measures that included use of marijuana, cocaine, crack, heroin, PCP, and ice.

3. 18-Month Follow-Up Characteristics and Wheels Task Performance. We now examine the relationship between characteristics of the mothers and children at the 18-month follow-up and their performance on the wheels task (see Table 5.11).

Demographic Characteristics. The age of the child is strongly associated with all measures on the wheels task. Older children named more wheeled objects correctly, named objects with fewer elicitations, and Mother's Ease of Ideas was higher with the older children. Older children have more extensive vocabularies than younger children as well as greater world knowledge so the ability to name wheeled objects is predictably related to the age of the child. As illustrated in the excerpts from transcripts presented earlier, older children are able to carry out the task with less prompting (fewer elicitations) from the mother. Finally, mothers may be more successful with older children than with younger children. Younger children have a smaller set of words for wheeled objects to retrieve, they may need more finely tuned clues, and mothers have to be more attentive to their knowledge of the children's experience in order to provide appropriate prompts.

Education and Literacy. Mothers with reading scores above the 7th grade level have children who name more wheeled objects and are rated higher on Ease of Ideas than those with lower reading scores at the 18-month follow-up.

Child Care. Those children who were in child care at some point during the 18-month follow-up period did better on Objects Named and Objects/Elicitations than those who were not. Children may be exposed to this type of guessing game in day care or preschool settings and able to transfer their facility to the interaction at home.

Psychosocial Characteristics. Child naming of wheeled objects and the Mother's Ease of Ideas were positively associated with the mothers' sense of mastery. Mothers with a greater sense of mastery over personal events may feel more confident in carrying out this task and more willing to elicit additional words from their children.

Contact with the fathers of the children was negatively associated with the Objects/Elicitations and Ease of Ideas. Mothers had lower scores if the fathers cared for the children overnight. As suggested above, the negative association between performance and contact with the fathers may be related to the children's age. Younger children have lower scores on these measures, and may be more likely to have contact with their fathers than the older children.

4. Evidence for Program Impacts on Mother-Child Interaction During the Wheels Task. There was no significant difference between mothers and children in the experimental group and those in the control group in the wheels task in terms of child performance (see Table 5.12). However, those in the control group received higher ratings on Mother's Ease of Ideas than those in the experimental group. Mothers in the experimental group had greater difficulty understanding the task and carrying it out successfully than those in the control group.

Examination of the distributions of scores in the two groups showed that a much higher proportion of mothers assigned to New Chance received the lowest rating on Mother's Ease of

Table 5.12

Impacts of New Chance on Wheels Task Variables

Observational Variable	Experimentals	Controls	Difference	p[a]
Objects Named	2.59	2.81	-0.22	0.430
Objects Named/Elicitations	0.26	0.26	0.00	0.449
Mother's Ease of Ideas	2.40	2.67	-0.27 **	0.012
Sample size	184	106		

Source: New Chance coded observational study variables.

Notes: Calculations for this table used data for all 290 respondents for whom there were observational data coded for literacy and 18-month follow-up survey data, including those with values of zero for outcomes and New Chance enrollees (i.e., experimentals) who did not participate in the program. The sample sizes may fall slightly short of the numbers reported because of videotape or audiotape problems.

The averages are adjusted using linear analysis of covariance procedures controlling for seven kinds of difference in characteristics before random assignment (age of child, maternal literacy, whether mother had more than one child, gender of child, race/ethnicity, Philadelphia site, and Portland site).

[a] An F-test was applied to each difference in regression-adjusted means. The column labeled "p" is the statistical significance level of each between-group impact. That is, p is the probability that the averages for the experimental and control groups are different from each other only because of random error. Statistical significance levels are indicated as ***<=1 percent; **<=5 percent; *<= 10 percent.

Ideas (23.4 versus 7.5 percent), a rating that indicates confusion about the expectations of the task or an inability to carry it out effectively. The proportion of mothers in the two groups scoring 2, 3, or 4 on this task were much less discrepant (22.8 percent of experimentals and 33.0 percent of controls scored 2, 42.9 and 45.5 percent scored 3, and 10.9 and 14.1 percent scored 4). A test of the difference in scores for those mothers in the experimental and control groups who *did* understand how to carry out the task (that is, who received any of the three higher ratings) indicated that these means did not differ significantly ($t = -.55$, $p = 0.55$). In other words, the difference between the experimentals and controls on the global rating Mother's Ease of Ideas can be attributed to those who received the lowest rating, that is, those who did not grasp the nature of the wheels task.

5. Program Impact for Subgroups. We now investigate the presence or absence of program impacts for subgroups on the Mother's Ease of Ideas variable (see Table 5.13).

There were no subgroups in which the mothers in the experimental group were rated significantly higher than those in the control group on Mother's Ease of Ideas. However, there were numerous subgroups in which mothers in the experimental group were rated significantly *lower* than those in the control group. These included mothers of male focal children, those with more than one child, those who were aged 18 and 19, those not living with their own mother or a partner, those who were aged 13–16 at the time of their first child's birth, and those who had a child under age 1 at baseline. In addition, mothers in the control group were rated higher than those in the experimental group in the subgroups who had completed only 10th grade or below, who had been out of school for 2 years or less, who scored at 9th grade or below on the reading measure, who had ever been employed, and who had always received welfare when they were growing up. Additionally, those mothers in the experimental group with a high risk for depression and those with lower sense of personal control were rated significantly lower than those in the control group on the wheels task. Most of these differences suggest that it was in somewhat more disadvantaged subgroups that experimental group mothers performed more poorly.

6. Summary of Results on the Wheels Task. The guessing game is unique among the mother-child tasks in this study in that it is a challenging, unstructured, open-ended, verbal game with no props. Successful strategies require generating a mental list of appropriate target words, isolating the most salient defining characteristics with which to prompt the child, taking into consideration the child's knowledge and vocabulary, and adjusting strategies according to the response of the child. Those who had the most difficulty with the wheels task were those with the fewest educational, social, and psychological resources (that is, low educational attainment, low literacy, high risk for depression). These mothers may have been particularly challenged to comprehend the intent of the task and stressed by the unstructured, test-like nature of the task and by its conceptual demands.

Among those with the fewest educational resources, the New Chance Program may have created a sense of test anxiety about this unstructured task, and resulted in a larger number of mothers in the experimental group missing the point of the task. Because we see no difference across groups among those who understood the task, we do not interpret the findings as pointing unambiguously to more or less effective performance among mothers in the experimental group.

Table 5.13

Impacts of New Chance on Ease of Ideas (Wheels Task) , by Subgroups Formed at Baseline

Characteristics and Subgroup at Baseline	n	Ease of Ideas		Within-Subgroup Impact	p^a	Between-Subgroups Difference	p^b
		Experimentals	Controls				
Full Sample							
Mean		2.40	2.67	-0.27 **	0.012		
Demographic Characteristics							
Child's age at observational study (months)						-----	0.821
24-36	86	2.41	2.72	-0.31	0.116		
37-47	115	2.40	2.58	-0.18	0.313		
48-63	89	2.41	2.73	-0.32	0.105		
Gender of child[c]						0.34	0.119
Female	142	2.40	2.50	-0.10	0.469		
Male	148	2.41	2.85	-0.44 ***	0.004		
Number of children						0.23	0.302
1	183	2.46	2.64	-0.18	0.195		
2 or more	107	2.29	2.70	-0.41 **	0.018		
Race/ethnicity						0.15	0.675
Black	244	2.38	2.63	-0.25	0.026		
White	46	2.56	2.96	-0.40 **	0.222		
Age of mother (years)						-----	0.352
16-17	56	2.46	2.63	-0.17	0.474		
18-19	148	2.29	2.71	-0.42 ***	0.006		
20-22	86	2.57	2.66	-0.09	0.673		
Living arrangement (mother)						0.20	0.389
Living with mother	104	2.42	2.57	-0.15	0.377		
Not living with mother	179	2.41	2.76	-0.35 **	0.013		
Living arrangement (partner/husband)						-0.20	0.572
Living with partner/husband	30	2.58	3.04	-0.46	0.194		
Not living with partner/husband	253	2.39	2.65	-0.26 **	0.028		
Age at first child's birth (years)						-0.26	0.230
13-16	135	2.34	2.75	-0.41	0.009		
17-19	155	2.45	2.60	-0.51 ***	0.338		
Age of youngest child (years)						-0.24	0.249
Less than one	107	2.13	2.56	-0.43	0.011		
One or older	183	2.55	2.74	-0.19 **	0.188		
Number of pregnancies						-0.03	0.923
1 or 2	218	2.44	2.71	-0.27	0.029		
3 or more	71	2.32	2.56	-0.24 **	0.252		
Education and Literacy							
Highest grade completed						-0.14	0.552
10th or below	185	2.34	2.66	-0.32	0.021		
11th or above	105	2.52	2.70	-0.18 **	0.288		

(continued)

Table 5.13 (continued)

Characteristics and Subgroup at Baseline	n	Ease of Ideas		Within-Subgroup Impact		p[a]	Between-Subgroups Difference		p[b]
		Experimentals	Controls						
Interval since last attended regular high school							-0.07		0.718
2 years or less	134	2.44	2.74	-0.30	**	0.050			
3 years or more	149	2.39	2.62	-0.23		0.121			
TABE reading test score[d]							-----		0.252
7th grade or below	132	2.28	2.62	-0.34	**	0.035			
8th or 9th grade	79	2.35	2.73	-0.38	*	0.068			
10th grade or above	79	2.70	2.67	0.03		0.858			
Employment and AFDC History									
Ever employed							-0.60	**	0.019
Yes	224	2.31	2.71	-0.40	***	0.001			
No	66	2.75	2.55	0.20		0.378			
Family received AFDC when sample member was growing up							-----		0.661
Always	60	2.31	2.76	-0.45	**	0.048			
Sometimes	145	2.40	2.62	-0.22		0.161			
Never	84	2.46	2.69	-0.23		0.273			
Psychosocial Characteristics									
CES-D Scale (depression) (%)[e]							-----		0.306
0-15 (not at risk)	141	2.47	2.59	-0.12		0.436			
16-23 (at some risk)	77	2.44	2.75	-0.31		0.136			
24-60 (at high risk)	72	2.24	2.76	-0.52	**	0.016			
Self-Esteem score[f]							0.60		0.585
Below the mean	75	2.35	2.72	-0.37	*	0.083			
At or above the mean (35)	215	2.43	2.66	-0.23	*	0.057			
Number of sources of emotional support							0.20		0.323
0-2	133	2.37	2.55	-0.18		0.257			
3 or more	157	2.43	2.81	-0.38	**	0.011			
Level of satisfaction with emotional support[g]							0.00		0.977
Less than very satisfied	138	2.42	2.69	-0.27	*	0.088			
Very satisfied	152	2.39	2.66	-0.27	*	0.068			
Locus of Control score[h]							0.57	**	0.011
Below the mean	115	2.20	2.82	-0.62	***	0.000			
At or above the mean (22)	175	2.53	2.58	-0.05		0.727			
Sample size		184	106						

Source and Notes: See Table 5.7.

Rather, we see reason for some concern that mothers in the New Chance experimental group, particularly those approaching it with the fewest strengths, seemed to be so unnerved by this task.[4] More in-depth qualitative analysis of these interactions, however, might help us understand better why this particular task is so challenging. An ongoing study of 48 mothers from this sample whose children were all aged 3 years plus or minus 5 months has found that almost 40 percent of these mothers misunderstood the task or showed no adjustment whatsoever to their children's confusion about it, but that these same mothers often showed considerable positive affect in their interactions over the task through tone of voice, facial expression, and gestures, indicating attention to and concern for their children's performance (Herot, 1996).

IV. Conclusions

In general, the pattern of results from the book reading task suggests that mothers in the New Chance Observational Study experimental group performed better both in overall Book Reading Quality and on some specific linguistic indicators of cognitively challenging interactions. These findings were most pronounced for those mothers with higher literacy scores and fewer indicators of social disadvantage. The patterns of adjustment to age of child shown in this sample mirrored the results of previous analyses, but even the experimental group mothers used less cognitively challenging talk than other samples of low-income mothers who have been studied (De Temple, 1994), confirming the high-risk social and academic status of the families involved in the New Chance Evaluation. The findings that children's experiences with fathers and with child care related to more positive performance in book reading highlight the value of *both* child-targeted and mother-targeted experiences in enhancing the quality of parent-child interaction.

There were no significant differences between experimental and control groups on the wheels task, except for the unexpectedly better performance of the control group on Mother's Ease of Ideas. For both the wheels and book reading tasks, children's experiences in child care settings related to enhanced performance, and mothers with higher levels of education and with higher literacy scores at follow-up also showed more sophisticated interactive strategies.

The most powerful predictor of good performance on both these tasks that are indicative of mothers' roles in educating and stimulating their children is their educational attainment. This finding is not surprising; it replicates not just previous findings concerning social class differences in North America, but also demographic studies in developing countries suggesting that educating women reduces their fertility and increases their likelihood of seeing themselves as having a role in educating their children. Important specifications of this relationship come from the finding that women who had achieved a regular high school diploma scored better than those who went back to school for a GED, and that literacy levels as measured by the TABE also relate strongly to task performance. These findings suggest, then, that programs that keep young women in school, even those who are starting their families at an early age, may well contribute to the quality of parenting and thus the school success of the next generation.

---

[4] As discussed in Part II of the monograph, this task also posed considerable challenges for interviewers, who were provided no tools for confirming the mother's comprehension and sometimes had difficulty clarifying its intent without either suggesting strategies or adding to the mother's confusion. Thus, it is fair to ask whether deficiencies on the part of the interviewers in presenting the task, possibly occurring with uneven frequency in the two groups, could have contributed to some mothers' difficulties. Based on the limited data we have on interviewer performance (see Part II), we find no direct evidence for this possibility. In only a handful of cases (2 percent of the sample) do we find a description of interviewer behavior on this task that could be characterized as "misleading or confusing presentation or interaction." This small number cannot account for the substantial number of mothers who failed to grasp the intent of the task or the magnitude of the difference between the two groups.

Chapter 6

Completing the Portrayal of Parenting Behavior with Interview-Based Measures

Donna R. Morrison, Martha J. Zaslow, and M. Robin Dion

To complete the portrayal of parenting for the observational study sample, Chapter 6 explores additional aspects of parenting gained in the context of interviews: both the 18-month follow-up and the brief interview carried out at the time of the observational session. In particular, we focus on an abbreviated version of the Home Observation for Measurement of the Environment (HOME) Inventory, which is designed to measure the cognitive stimulation and emotional support available to children in the home setting; three Maternal Report Parenting scales, which provide mothers' subjective perspectives on their parenting experiences; and two Time Use measures, which reveal how much time mothers allocated to parenting activities. Analyses of these interview-based measures for the observational study sample reveal positive impacts of the New Chance Program on both the quantity and quality of parenting. On the Time Use measures, experimental group mothers reported spending more time on parenting activities overall, and on parenting chores in particular, such as bathing and feeding their children. On the abbreviated HOME Inventory, the home environment was rated as more supportive and stimulating overall for mothers in the experimental group, and a difference was found particularly in the area of the emotional support available to the focal child in the home. The findings on the HOME Inventory are in accord with the self-perceptions of the mothers as reflected in the Maternal Report Parenting scale focusing on warmth: experimental group mothers perceived themselves as showing greater warmth to the focal child.

Although the primary focus of this monograph is on *observational* measures of mother-child interaction, additional insight may be gained by examining measures of parenting derived in the context of interviews carried out with study participants. We refer to the combination of *maternal report measures of parenting* collected in the context of interviews, and *interviewer ratings* of the home environment and of the mother-child relationship, as "interview-based measures of parenting." Such measures were the means through which program impacts on parenting behavior were assessed in the 18-month follow-up of the full New Chance Evaluation. These measures are thus available for the embedded study sample as well. In addition, in a brief interview with the mothers in the embedded study that was carried out at the time of the observational session (the "brief interview"), further interview-based measures of parenting were obtained.

Findings for the interview-based measures of parenting available for the embedded study sample, combined with those for the observational measures, yield a comprehensive view of parenting behavior for this sample. Further, it is important to ask whether findings for the observational and interview-based measures of parenting are in accord within the observational study sample or whether measures obtained in these different ways yield a divergent picture of program impacts.

Accordingly, in this chapter, we complete the picture of parenting behavior for the observational study sample by focusing on interview-based measures of parenting. As with the observational measures, we progress from a presentation of descriptive findings (asking how the interview-based measures of parenting vary in light of family background characteristics for the sample as a whole) to an examination of program impacts (examining experimental-control group differences overall as well as for key subgroups).

Three sets of interview-based measures of parenting are available for the New Chance Observational Study sample: (1) four subscales and a total score from an abbreviated version of the HOME Inventory (HOME-SF), collected as part of the full evaluation at the 18-month follow-up survey; (2) three Maternal Report Parenting scales, providing the mother's subjective perspective on the experience of parenting, also collected as part of the full evaluation at the 18-month follow-up; and (3) two parenting Time Use measures, collected only for the observational study sample as part of the brief interview at the time of the observational session.

These measures differ from, and complement, the observational measures of parenting in several important respects (see also Chapter 7). For example, the observational measures focus exclusively on interactions within the mother-child dyad. Further, they do so with the situation and materials held constant (to the extent feasible) for all families, focusing on a context in which the mother is asked to be a resource to her child during specific challenging tasks. By contrast, the interview-based measures seek to document behavior (and feelings about behavior) not in a specified context but in the multiple contexts of everyday life. Thus, rather than seeking to "even the playing field" in terms of the materials or time available to mothers, the interview-based measures seek to document how materials and time typically serve as resources for mother-child interaction, as well as the nature of these interactions. While the observational measures provide extremely sensitive qualitative measures of interaction, they do not attempt to provide the mother's subjective sense of her relationship with the child or her experience of parenting. The interview-based measures include (although they are not restricted to) the mother's perspective.[1] Furthermore, while the observational measures focus only on the mother-child dyad, the interview-based measures consider the input of other social partners in the home, for example, other family members who may provide cognitive stimulation by taking the child on outings. Finally, the informant, or source of data, for the observational measures is a highly trained behavioral coder. By contrast, the interview-based measures of parenting rely on information from the mother and an interviewer.[2]

Because of these differences in the nature and focus of the observational and interview-

---

[1] The abbreviated HOME Inventory is based on both maternal report and interviewer rating, and thus includes both the mother's and the interviewer's perspectives. The remaining interview-based measures included in this study rely entirely on maternal report.

[2] See also the discussion in Chapter 10 concerning the issue of consistent method effects when both predictor and outcome variables rely on a single source.

BEST COPY AVAILABLE

based measures of parenting, the differing parenting measures can be seen as *jointly* providing a comprehensive view of parenting behavior and the home environment. It is indeed interesting to note that the complementary nature of observational and interview-based measures is implicitly acknowledged by other researchers through the inclusion of both types of measures in studies of parenting behavior (see, for example, Berlin et al., 1995, regarding contrasting measures of parenting used in the evaluation of the Infant Health and Development Program; among studies of adolescent mothers in particular, see Crockenberg, 1987; Garcia-Coll, Hoffman, and Oh, 1987; Schilmoeller and Baranowski, 1985; Unger and Wandersman, 1988; and Walker et al., 1995).

In Chapter 7 we examine the full set of parenting measures available for the embedded study sample, comparing the magnitude of program impacts on the various measures through an examination of effect sizes. We use a common analytic strategy to explore the extent to which participation in various components of the New Chance Program helps to explain program impacts on a number of parenting behaviors. Finally, we examine how closely and in what patterns the various measures are correlated.

Before taking this look *across* all of the parenting measures, we document the pattern of findings, within the observational study sample, for the interview-based measures of parenting. We begin by briefly describing each of the interview-based measures.

I. A Description of the Interview-Based Measures of Parenting

A. The Abbreviated HOME Inventory

The original Home Observation for Measurement of the Environment (HOME) Inventory, from which a short form (HOME-SF) was later derived for survey administration, was developed by Caldwell and Bradley (1984) and is a widely used measure of the cognitive stimulation and emotional support available to children in the home setting. Separate versions exist to describe the home environments of infants and toddlers, preschool-age children, and school-age children.

It is important to note that the original HOME Inventory was conceptualized as a summary index of features of the physical and interpersonal environment of the home that pose risk to children's development (Elardo and Bradley, 1981). Thus, based on a home visit during which the mother and a particular child are present, questions answered by the mother and ratings made by the visitor are each summarized as dichotomous variables, indicating whether input to the child was or was not in a range involving risk to development. Unlike other parenting measures, the HOME Inventory does not provide a description of the home environment encompassing a full range, from least to most, of possible interpersonal or environmental inputs to the child. Rather this measure truncates the potential range for each measure in order to portray the presence or absence of risk. Thus, while data from the observational measures provides a full portrayal of the literacy-related and affective quality of mother-child interactions, the HOME Inventory provides summary measures of risk for different aspects of the child's home environment.

The HOME Inventory has an impressive track record in terms of concurrent and predictive validity (evidence recently summarized and evaluated by Mariner and Zaslow, 1997). This measure has been found to be related to such important characteristics of the family as mother's and father's education, father's presence, father's occupation, crowding, family and neighborhood poverty, overall family socioeconomic status, and maternal depression (see, for example, Affleck et al., 1982; Bradley and Caldwell, 1984; Bradley, Caldwell, and Rock, 1988; Hollenbeck, 1978; Klebanov, Brooks-Gunn, and Duncan, 1994; Kurtz, Borkowski, and Deshmukh, 1986; Parks and Bradley, 1991). HOME total and specific subscale scores have also been found to be predictive of child outcomes and especially measures of cognitive development. For example, studies report the HOME scales to be significant predictors of child IQ and academic achievement (Bradley and Caldwell, 1979, 1980, and 1984; Bradley, Caldwell, and Rock, 1988; Bradley et al., 1989; Ramey, Yates, and Short, 1984). There are also reports of significant prediction from the HOME total and subscale scores to measures of socioemotional development (for example, the occurrence of behavior problems), although prediction in this domain of development is not as strong (Berlin et al., 1995; Parks and Bradley, 1991).

The originators of the HOME Inventory assisted in the development of a shortened form of this measure (HOME-SF) for use in the National Longitudinal Survey of Youth Child Supplement (Baker and Mott, 1989). The HOME-SF includes about half of the items of the full form and involves a reformatting of the mother-interviewer interaction into a standardized rather than open-ended interview, as well as a rewording of some items. New ground has been broken with the inclusion of this measure of the home environment, together with child assessments, in the context of a national survey with extensive information about the mothers and families: the National Longitudinal Survey of Youth Child Supplement (Chase-Lansdale et al., 1991). While the concurrent and predictive validity of the HOME-SF has been less extensively examined than that of the full HOME Inventory, the evidence available to date indicates that the short form also predicts to both cognitive and socioemotional child outcomes (see, for example, Baker et al., 1993; Desai, Michael, and Chase-Lansdale, 1989; Moore and Snyder, 1991; Sugland et al., 1995).

The New Chance 18-month follow-up included the HOME-SF, with the infant and toddler, preschool-age, or school-age version used according to the age of the focal child. While the National Longitudinal Survey of Youth-Child Supplement provides a total score and two subscale scores (Emotional Support and Cognitive Stimulation), the 18-month follow-up of the New Chance Evaluation instead summarizes the HOME-SF in terms of a total score and four subscales: Emotional Support, Cognitive Stimulation, Physical Environment, and Harsh Discipline. Because of low internal consistency for these subscales when individual items were summarized into the dichotomous variables, scoring for the New Chance 18-month follow-up preserved the trichotomous nature of the items as worded in the interview. To permit comparison across the three versions of the HOME-SF used for different age groups, standardized summary scores were created with a mean of 100 and a standard deviation of 15 (see Quint et al., 1994, for a more detailed discussion). It is important to note that unlike the other two measures of discipline in this study (Maternal Control/Punitiveness and the observational measure of Harsh Treatment) the direction of scoring for the Harsh Discipline subscale of the HOME-SF as used here is such that a higher score indicates *less* harsh discipline. In this way, higher scores on all HOME subscales indicate better parenting.

In the full New Chance Evaluation sample, internal consistency scores for the HOME-SF total and the four subscales were computed separately for the infant and toddler, preschool-age and school-age versions. For the HOME-SF total, internal consistency ranged from .70 to .76 depending on age group. For the four HOME subscales, internal consistency for the different age groups ranged from .44 to .53 for Cognitive Stimulation, from .57 to .60 for Emotional Support, from .71 to .82 for Physical Environment, and from .41 to .49 for Harsh Discipline. In the full evaluation sample, a small but significant difference was found on the HOME Emotional Support subscale, with experimental group mothers providing more emotional support to their young children.

B. The Maternal Report Parenting Scales

In the full New Chance Evaluation, three Maternal Report Parenting scales were included in order to address specific research concerns or to supplement the information available through the HOME-SF.

A Parenting Stress scale was included in the 18-month New Chance follow-up to tap the mothers' perceptions of stress and aggravation in the parenting role. It is important to note that this scale explores the mothers' *subjective* sense of difficulty specifically with regard to the parenting role, rather than stress from other aspects of their lives.

Within the New Chance sample, an important question is whether participation in parenting education or in other parts of the intervention reduce parenting stress. Stress may diminish if changes in parenting behavior occur that involve an increase in maternal effectiveness — for example, in dealing with disobedience in the child. Stress may also diminish in the absence of changes in parenting behavior — for example, if mothers come to view the problems they are experiencing in the parenting role as more universal. Thus, it is of interest to examine program impacts on mothers' subjective sense of the parenting role whether or not there are measurable program impacts on mother-child interaction or the home environment.

The Parenting Stress scale drew upon, but modified, items from the Parenting Stress Index (Abidin, 1983) as well as a number of other sources. This 8-item scale had an internal consistency of .70 in the full evaluation sample. Scores have a potential range from 0 (a complete absence of parenting stress) to 80 (very substantial parenting stress). Within the full evaluation sample, mothers reported fairly low levels of parenting stress (mean scores of 27.6 and 28.2 in the experimental and control groups, respectively, with no significant group difference).

A brief (3-item) Maternal Warmth/Responsiveness scale was included in the 18-month follow-up because the HOME-SF relies relatively heavily on interviewer ratings for the Emotional Support subscale, while other HOME-SF subscales (for example, the Cognitive Stimulation subscale) rely fairly equally on maternal report and interviewer rating items. The Maternal Warmth/Responsiveness scale makes it possible to ask whether differences in this dimension of parenting behavior also occur in the eyes of the mother.

Maternal warmth (both alone and in combination with disciplinary approach) has been identified as an important aspect of the parent-child relationship. For example, greater maternal warmth predicts to teacher ratings of fewer behavior problems and greater social competence in the early years of elementary school (Patterson, Cohn, and Kao, 1989). Warmth in the mother-child relationship is an important protective factor, diminishing the likelihood of psychopathology later in development (see, for example, Rutter, 1987).

Internal consistency for the Maternal Warmth scale in the full New Chance Evaluation sample was .55. Scores on this scale have a potential range from 0 (very low self-perception of warmth) to 30 (very high). The mean for this scale in the full evaluation sample was 23.5, indicating that mothers perceive themselves highly favorably in terms of warmth; and no group difference between the experimental and control groups was detected on this variable.

The Maternal Control/Punitiveness scale included in the 18-month follow-up in the full evaluation goes beyond the HOME Harsh Discipline subscale by addressing an important aspect of discipline other than the use of such strategies as physical punishment: that is, whether the mother tends to use an authoritarian approach toward the child, expecting obedience irrespective of the circumstances, or tends to use reasoning and explanation.

Coercive, as opposed to inductive, control strategies predict measures of child social competence with peers (see, for example, Dishion, 1990; Hart, Ladd, and Burleson, 1990). Outcomes for children differ especially when the dimensions of disciplinary approach and warmth/positive involvement are considered simultaneously (Maccoby and Martin, 1983). The combination of firm but reasoning-based discipline with high support and nurturance, termed an "authoritative" parenting style, is predictive of positive child outcomes both in early development and through adolescence. By contrast, an "authoritarian" parenting style, involving high control but limited warmth, is predictive of less positive development (see, for example, Baumrind, 1971; Steinberg, 1990).

The 6-item Maternal Control/Punitiveness scale in the New Chance Evaluation was intended to go beyond the issue of physical punishment to examine the mother's use of power-assertive strategies as opposed to democratic disciplinary strategies (Quint et al., 1994). Internal consistency on this scale in the full evaluation sample at 18 months was .60. Scores on this scale have a potential range from 0 (extremely low control and punitiveness) to 60 (extremely high control and punitiveness). Within the full evaluation sample, mothers in the experimental group reported significantly less authoritarian control than those in the control group (mean scores for the experimental and control groups were 21.3 and 23.7, respectively).

C. Time Use Measures

There are notable variations in the amount of time mothers spend with their young children. Studies link maternal time use with children to maternal background characteristics. For example, one study revealed that more economically advantaged mothers allocate substantially more of their nonmarket-labor time to their preschool-age children than lower-income mothers (Hill and Stafford, 1974). Further, time use research suggests that "time parents spend with children has implications for their developmental and educational outcomes" (Bryant and Zick, 1993, p. 1).

Just as the New Chance Intervention may alter mothers' subjective perception of stress in the parenting role, with or without change in parenting behavior, participation in New Chance may also alter allocation of time to the parenting role, with or without changing the nature or quality of interactions. For example, participation may result in improvement in the quality of mother-child interaction, but a diminution in the time that mothers spend with their young children because of educational or employment commitments. Alternately, participation in New Chance may succeed in encouraging mothers to devote more time to the parenting role without resulting in an enhanced quality of mother-child interaction. In order to understand the effects of New Chance on parenting, it is important to consider both time allocated to this role by the young mothers and quality of the time spent together.

Accordingly, as part of the brief interview completed with mothers in the New Chance Observational Study, they were asked to describe their activities and those of the focal child during the most recent weekday (a procedure developed by Polit and Eldred specifically for the New Chance Observational Study). For each type of activity the interviewer inquired "what time?" and "how long?" The question was open-ended. Each mother described her day, with the interviewer probing to elicit her activities, broken down into half-hour intervals. The responses were later coded into specific behavior categories (derived by Polit and Eldred).

Two of the behavior categories are specifically indicative of parenting time: (1) a summary of total time spent on parenting activities (Overall Parenting Time), and (2) a summary of time spent on parenting chores, such as dressing, bathing, or feeding the child (Parenting Chore Time). Whereas the other interview-based measures of parenting ask the mother to consider her relationship and behavior with the focal child only, the Time Use measures ask about time spent with any of her children.[1]

There was a wide range in the number of hours mothers reported spending on parenting during the previous 24-hour period: from zero to 11 hours on all parenting activities combined, and from zero to 9 hours on parenting chores specifically. However, the *average* amount of time that mothers reported spending on all parenting activities was only 2 1/4 hours. The average time that mothers reported spending on parenting chores specifically was only a little over an hour.

There are no comparable Time Use data for the full evaluation sample, as this information was collected only as part of the brief interview at the time of the observational session.

II. Variation in the Interview-Based Measures of Parenting According to Baseline and 18-Month Characteristics

A. Baseline Subgroup Scores on the Interview-Based Measures of Parenting

We turn now to the question of whether there was variation in each of the interview-based measures of parenting according to characteristics of the mothers, children, and families at the time of enrollment in the New Chance Evaluation.

---

[1] Parenting time may be related to the number of children in a family. We have noted that mothers in the control group within the observational study sample tended to have more children at baseline (a difference that approached significance). It is therefore important to note that all impact analyses control for number of children in the family at baseline. Thus, any differences in time allocated to parenting on the part of experimentals versus controls are net of differences at baseline in family size.

Table 6.1 shows mean HOME-SF scores for subgroups defined according to characteristics at baseline. Tables 6.2 and 6.3 present means for the Maternal Report Parenting scales and Time Use measures for the same subgroups. Means in these tables have been adjusted for experimental/control group status in order to be descriptive of the overall sample. That is, in order to discuss the sample as a whole, we have statistically removed any effect of experimental group status in these tables.

Our summary of findings below (as in previous reporting for the full New Chance Evaluation) reports on associations between *individual* baseline and 18-month characteristics and *individual* parenting scales. In interpreting the results, it should be kept in mind that the individual baseline and 18-month measures to which we are relating the interview-based measures of parenting may, in fact, be correlated. For example, mothers' education and literacy may be closely associated; mothers' pattern of residence and number of children may be linked; mother's race/ethnicity may be correlated with use of child care. In a similar way, the discrete parenting variables considered here may also be intercorrelated. The possibility of such associations across variables should qualify our interpretation of findings. That is, any significant correlation between a single family characteristic and a single parenting variable may be a reflection of a more complex set of relationships. A further concern with the presentation of findings in terms of discrete variables is the increased possibility of chance findings.

Accordingly, in our presentation of descriptive results below, we report differences on parenting measures in light of a particular subgroup variable (such as mother's education at baseline), only when statistically significant differences were observed for that variable on at least two of the interview-based measures of parenting (for example, if differences had occurred in light of maternal education on both the HOME-SF Cognitive Stimulation subscale, *and* the Maternal Control scale).

Demographic Characteristics. Significant patterns of association were observed between several of the interview-based measures of parenting (particularly the HOME-SF) and demographic characteristics of mothers such as number of children, race, age, and timing of first birth.

Having fewer children at baseline was associated with more favorable HOME-SF total scores, as well as Cognitive Stimulation and Physical Environment subscale scores at the 18-month follow-up among mothers in the New Chance Observational Study. Women with one child had HOME-SF total scores roughly 5 points higher on average than those with more than one child. In terms of *quantity* of time, mothers with more than one child spent slightly more time on parenting chores than their counterparts with a single child. It is not clear whether differences in the quality of the home environments provided in only-child versus multichild families is attributable to the relatively undivided attention of mothers with one child or to underlying characteristics of women who choose to limit or space their fertility. These same factors (for example, high expectations for their children's achievement) could also affect the investments that mothers make in the children already born.

White mothers scored less favorably than black mothers on the Harsh Discipline subscale of the HOME-SF and rated themselves as notably higher on power-assertive disciplinary styles

Table 6.1

Means on HOME-SF Scales, by Subgroups Formed at Baseline

Subgroup Formed at Baseline	n	HOME-SF Emotional Support	HOME-SF Cognitive Stimulation	HOME-SF Physical Environment	HOME-SF Harsh Discipline	HOME-SF Total
Full Sample						
Mean		96.68	98.36	97.61	97.92	96.30
Standard deviation		16.07	15.20	16.70	15.40	16.03
Range		50-127	58-135	51-118	39-119	45-129
Demographic Characteristics						
Gender of child[a]						
Male	139	96.58	99.59 *	97.68	97.40	97.03
Female	139	96.84	95.70	96.73	98.58	95.20
Number of children						
1	177	96.76	99.25 *	98.98 *	98.69	97.78 *
2 or more	101	96.63	94.89	94.16	96.78	93.27
Race/ethnicity						
White	233	97.41	97.04	97.58	97.24 *	96.08
Black	45	93.08	100.76	95.24	101.85	96.30
Age of mother (years)						
16-17	54	92.30 *	92.29 *	95.02	97.37 *	90.58 *
18-19	141	98.48	98.76	96.30	99.94	97.35
20-22	83	96.51	99.19	100.17	95.01	97.57
Living arrangement (mother)						
Living with mother	103	97.49	97.35	97.17	98.43	96.32
Not living with mother	168	96.58	97.85	97.21	97.64	96.16
Living arrangement (partner/husband)						
Living with partner/husband	29	89.81 **	97.50	91.53 *	96.99	90.41 *
Not living with partner/husband	242	97.78	97.68	97.89	98.05	96.92
Age at first child's birth (years)						
13-16	129	97.34	95.81 *	95.66	98.37	94.90
17-19	149	96.18	99.18	98.49	97.66	97.14
Age of youngest child (years)						
Under 1	103	96.15	95.94	94.52 *	98.70	94.08
1 or older	175	97.04	98.67	98.82	97.55	97.34
Number of pregnancies						
1 or 2	208	95.94	98.23	97.24	98.92	96.32
3 or more	69	99.52	96.15	97.27	95.87	96.10

(continued)

Table 6.1 (continued)

Subgroup Formed at Baseline	n	HOME-SF Emotional Support	HOME-SF Cognitive Stimulation	HOME-SF Physical Environment	HOME-SF Harsh Discipline	HOME-SF Total
Education and Literacy						
Highest grade completed						
10th or below	178	94.49 **	95.00 ***	95.26 **	97.99	93.19 ***
11th or above	100	100.63	102.32	100.64	97.97	101.31
Interval since last attended regular high school						
2 years or less	131	96.81	98.85	97.18	99.61 *	96.98
3 years or more	140	96.57	97.03	97.11	96.33	95.45
TABE reading test score (grade equivalent)[b]						
7th grade or below	134	94.85 *	96.74	96.00	97.29	94.22
8th or 9th grade	67	100.11	96.11	100.26	97.52	97.92
10th grade or above	75	96.58	100.63	96.30	99.60	97.59
Employment and AFDC History						
Ever employed						
No	64	94.43	96.61	92.37 *	99.76	92.79 *
Yes	214	97.36	97.94	98.58	97.48	97.07
Family received AFDC when sample member was growing up						
Never	79	98.54	98.84	98.97	96.70	98.00
Sometimes	138	96.22	96.85	96.40	97.67	94.98
Always	60	95.59	97.78	96.35	101.15	96.26
Psychosocial Characteristics						
CES-D (depression) score[c]						
0-15 (not at risk)	134	96.63 *	100.05 *	97.33	100.60 *	97.98 *
16-23 (at some risk)	74	99.58	96.77	98.91	96.75	97.10
24-60 (at high risk)	70	93.78	93.90	95.12	94.21	91.42
Self-Esteem score[d]						
Below mean	73	93.52	92.09	96.86	97.17	91.81
At or above mean (35)	205	97.78 *	99.52 ***	97.32	98.26	97.57 ***
Number of sources of emotional support						
0-2	125	96.46	97.30	97.87	97.83	96.07
3 or more	153	96.91	97.92	96.67	98.11	96.15

(continued)

Table 6.1 (continued)

Subgroup Formed at Baseline	n	HOME-SF Emotional Support	HOME-SF Cognitive Stimulation	HOME-SF Physical Environment	HOME-SF Harsh Discipline	HOME-SF Total
Level of satisfaction with emotional support[e]						
Less than very satisfied	131	96.78	97.96	95.51	98.13	95.67
Very satisfied	147	96.65	97.38	98.66	97.86	96.50
Locus of Control score[f]						
Below mean	112	96.22	94.31 **	97.16	96.65	93.80 *
At or above mean (22)	166	97.04	99.89	97.23	98.88	97.68
Sample size	290					

Sources: New Chance baseline enrollment data and 18-month follow-up survey.

Notes: Means are adjusted for experimental/control status in order to be descriptive of the overall sample.

Statistical significance levels are indicated as *** < = 1 percent, ** < = 5 percent, and * < = 10 percent, which refers to the probability that sample means are different from each other only because of random error.

For subgroups that have more than two levels, the placement of asterisks does not indicate where the significant difference lies, but only that there was a significant difference among the set of means.

Calculations for this table used data for all 290 respondents for whom there were 18-month follow-up survey data, including those with values of zero for outcomes and New Chance enrollees (i.e., experimentals) who did not participate in the program. The sample size may fall slightly short of the number reported because of missing or unusable items from some respondents' questionnaires.

Direction of scoring was reversed for the Harsh Discipline scale, so that a higher score indicates better (less harsh) parenting behavior.

[a] When a sample member had more than one child, her response refers to the randomly selected focal child.

[b] The test used to measure reading ability was the reading part of the Tests of Adult Basic Education (TABE). Most sites administered the survey form of the test, but some administered the full reading test.

[c] The Center for Epidemiological Studies Depression (CES-D) Scale is a widely used measure of depression; scores can range from zero to 60.

[d] The measure of self-esteem used was the Rosenberg Self-Esteem Scale, a 10-item scale that assesses a person's global sense of self-worth. Scores can range from 10 to 50; 30 is considered the neutral midpoint.

[e] Enrollees were asked about their degree of satisfaction with the emotional support ("people who listen to you, reassure you, and show you they care") they received. Levels range from 1 (very dissatisfied) to 5 (very satisfied); however, because so few individuals indicated that they were less than very satisfied with their emotional support, levels 1 through 4 were collapsed into one category labeled "less than very satisfied."

[f] The Locus of Control Scale is a 6-item adaptation of the longer scale originally developed by Julien Rotter (1966). Scores can range from 6 to 30; 18 is considered the neutral midpoint.

Table 6.2

Means on Maternal Report Parenting Scales, by Subgroups Formed at Baseline

Subgroup Formed at Baseline	n	Parenting Stress	Maternal Warmth	Maternal Control/ Punitiveness
Full Sample				
Mean		23.60	29.61	23.94
Standard deviation		6.28	14.52	11.96
Range		0-30	0-68	0-60
Demographic Characteristics				
Gender of child[a]				
Male	141	29.01	23.74	23.87
Female	146	30.82	23.42	24.53
Number of children				
1	181	28.87	23.27	23.63
2 or more	106	31.71	24.10	25.17
Race/ethnicity				
White	45	30.17	23.84	25.62 ***
Black	242	28.56	22.21	16.81
Age of mother (years)				
16-17	55	31.00	24.79	25.25
18-19	147	28.84	23.30	22.75
20-22	85	31.06	23.27	26.01
Living arrangement (mother)				
Living with mother	103	29.86	23.60	25.20
Not living with mother	177	29.57	23.70	23.90
Living arrangement (partner/husband)				
Living with partner/husband	29	27.48	22.24	18.89 **
Not living with partner/husband	251	29.95	23.84	25.05
Age at first child's birth (years)				
13-16	133	30.26	24.37 *	24.47
17-19	154	29.62	22.92	23.97
Age of youngest child (years)				
Under 1	106	30.59	23.59	23.73
1 or older	181	29.50	23.57	24.48
Number of pregnancies				
1 or 2	216	29.32	23.59	23.89
3 or more	70	31.53	23.70	25.33

(continued)

Table 6.2 (continued)

Subgroup Formed at Baseline	n	Parenting Stress	Maternal Warmth	Maternal Control/ Punitiveness
Education and Literacy				
Highest grade completed				
10th or below	182	30.71	23.61	23.98
11th or above	105	28.50	23.52	24.58
Interval since last attended regular high school				
2 years or less	133	29.56	24.12	23.94
3 years or more	148	30.06	23.25	24.35
TABE reading test score (grade equivalent)[b]				
7th grade or below	135	31.27	23.66	27.57 ***
8th or 9th grade	70	29.92	23.68	23.48
10th grade or above	80	27.62	23.35	19.20
Employment and AFDC History				
Ever employed				
No	64	30.63	24.52	25.08
Yes	223	29.71	23.31	23.94
Family received AFDC when sample member was growing up				
Never	83	31.85	22.66	23.15
Sometimes	145	29.98	24.05	24.38
Always	58	27.65	23.77	25.60
Psychosocial Characteristics				
CES-D (depression) score[c]				
0-15 (not at risk)	139	26.76 **	23.88	22.63
16-23 (at some risk)	77	30.37	22.95	25.38
24-60 (at high risk)	71	35.59	23.68	25.97
Self-Esteem score[d]				
Below mean	74	36.19	23.56	27.18
At or above mean (35)	213	27.80 **	23.59	23.19
Number of sources of emotional support				
0-2	132	30.62	23.12	24.35
3 or more	155	29.35	23.94	24.07
Level of satisfaction with emotional support[e]				
Less than very satisfied	137	29.73	22.87 *	23.94
Very satisfied	173	30.07	24.19	24.42

(continued)

Table 6.2 (continued)

Subgroup Formed at Baseline	n	Parenting Stress	Maternal Warmth	Maternal Control/ Punitiveness
Locus of Control score[f]				
Below mean	114	33.80 **	23.58	26.49 **
At or above mean (22)	173	27.30	23.58	22.65
Sample size	290			

Sources: New Chance baseline enrollment data and 18-month follow-up survey.

Notes: Means are adjusted for experimental/control status in order to be descriptive of the overall sample.

Statistical significance levels are indicated as *** < = 1 percent, ** < = 5 percent, and * < = 10 percent, which refers to the probability that sample means are different from each other only because of random error.

For subgroups that have more than two levels, the placement of asterisks does not indicate where the significant difference lies, but only that there was a significant difference among the set of means.

Calculations for this table used data for all 290 respondents for whom there were 18-month follow-up survey data, including those with values of zero for outcomes and New Chance enrollees (i.e., experimentals) who did not participate in the program. The sample size may fall slightly short of the number reported because of missing or unusable items from some respondents' questionnaires.

[a] When a sample member had more than one child, her response refers to the randomly selected focal child.

[b] The test used to measure reading ability was the reading part of the Tests of Adult Basic Education (TABE). Most sites administered the survey form of the test, but some administered the full reading test.

[c] The Center for Epidemiological Studies Depression (CES-D) Scale is a widely used measure of depression; scores can range from zero to 60.

[d] The measure of self-esteem used was the Rosenberg Self-Esteem Scale, a 10-item scale that assesses a person's global sense of self-worth. Scores can range from 10 to 50; 30 is considered the neutral midpoint.

[e] Enrollees were asked about their degree of satisfaction with the emotional support ("people who listen to you, reassure you, and show you they care") they received. Levels range from 1 (very dissatisfied) to 5 (very satisfied); however, because so few individuals indicated that they were less than very satisfied with their emotional support, levels 1 through 4 were collapsed into one category labeled "less than very satisfied."

[f] The Locus of Control Scale is a 6-item adaptation of the longer scale originally developed by Julien Rotter (1966). Scores can range from 6 to 30; 18 is considered the neutral midpoint.

Table 6.3

Means on Parenting Time Use Measures, by Subgroups Formed at Baseline

Subgroup Formed at Baseline	n	Overall Parenting Time	Parenting Chore Time	
Full Sample				
Mean		2.25	1.15	
Standard deviation		2.23	1.23	
Range		0-11	0-9	
Demographic Characteristics				
Gender of child[a]				
Female	142	2.20	1.14	
Male	148	2.28	1.16	
Number of children				
1	183	2.16	1.03	**
2 or more	107	2.39	1.36	
Race/ethnicity				
Black	244	2.28	1.14	
White	46	2.12	1.18	
Age of mother (years)				
16-17	56	2.02	1.05	
18-19	148	2.11	1.05	
20-22	86	2.62	1.38	
Living arrangement (mother)				
Living with mother	104	2.19	1.18	
Not living with mother	179	2.22	1.06	
Living arrangement (partner/husband)				
Living with partner/husband	30	2.05	0.98	
Not living with partner/husband	253	2.23	1.12	
Age at first child's birth (years)				
13-16	135	2.14	1.15	
17-19	155	2.33	1.15	
Age of youngest child (years)				
Under 1	107	2.01	1.09	
1 or older	183	2.38	1.18	
Number of pregnancies				
1 or 2	218	2.17	1.11	
3 or more	71	2.50	1.29	

(continued)

Table 6.3 (continued)

Subgroup Formed at Baseline	n	Overall Parenting Time	Parenting Chore Time
Education and Literacy			
Highest grade completed			
10th or below	185	2.27	1.08
11th or above	105	2.21	1.27
TABE reading test score (grade equivalent)[b]			
7th grade or below	132	2.23	1.14
8th or 9th grade	79	2.23	1.22
10th grade or above	79	2.28	1.08
Employment and AFDC History			
Ever employed			
Yes	224	2.26	1.19
No	66	2.18	1.02
Family received AFDC when sample member was growing up			
Never	84	2.36	1.24
Sometimes	145	2.22	1.17
Always	60	2.17	0.99
Psychosocial Characteristics			
CES-D (depression) score[c]			
0-15 (not at risk)	141	2.32	1.10
16-23 (at some risk)	77	1.92	1.08
24-60 (at high risk)	72	2.44	1.31
Self-Esteem score[d]			
Below mean	75	2.30	1.30
At or above mean (35)	215	2.23	1.10
Number of sources of emotional support			
0-2	133	2.12	1.16
3 or more	157	2.35	1.14
Level of satisfaction with emotional support[e]			
Less than very satisfied	138	2.49	1.24
Very satisfied	152	2.02	1.06
Locus of Control score[f]			
Below mean	115	2.32	1.18
At or above mean (22)	175	2.19	1.13
Sample size	290		

(continued)

Table 6.3 (continued)

Sources: New Chance baseline enrollment data and data from the brief interview accompanying the observational session.

Notes: Means are adjusted for experimental/control status in order to be descriptive of the overall sample.

Statistical significance levels are indicated as *** < = 1 percent, ** < = 5 percent, and * < = 10 percent, which refers to the probability that sample means are different from each other only because of random error.

For subgroups that have more than two levels, the placement of asterisks does not indicate where the significant difference lies, but only that there was a significant difference among the set of means.

Calculations for this table used data for all 290 respondents for whom there were 18-month follow-up survey data, and data from the brief interview accompanying the observational session, including those with values of zero for outcomes and New Chance enrollees (i.e., experimentals) who did not participate in the program. The sample size may fall slightly short of the number reported because of missing or unusable items from some respondents' questionnaires.

[a] When a sample member had more than one child, her response refers to the randomly selected focal child.

[b] The test used to measure reading ability was the reading part of the Tests of Adult Basic Education (TABE). Most sites administered the survey form of the test, but some administered the full reading test.

[c] The Center for Epidemiological Studies Depression (CES-D) Scale is a widely used measure of depression; scores can range from zero to 60.

[d] The measure of self-esteem used was the Rosenberg Self-Esteem Scale, a 10-item scale that assesses a person's global sense of self-worth. Scores can range from 10 to 50; 30 is considered the neutral midpoint.

[e] Enrollees were asked about their degree of satisfaction with the emotional support ("people who listen to you, reassure you, and show you they care") they received. Levels range from 1 (very dissatisfied) to 5 (very satisfied); however, because so few individuals indicated that they were less than very satisfied with their emotional support, levels 1 through 4 were collapsed into one category labeled "less than very satisfied."

[f] The Locus of Control Scale is a 6-item adaptation of the longer scale originally developed by Julien Rotter (1966). Scores can range from 6 to 30; 18 is considered the neutral

(Maternal Control/Punitiveness) than black mothers (scores of 25.6 and 16.8, respectively). It is difficult to hypothesize why these patterns across racial groups emerged for these variables.

Comparatively older mothers provided more favorable home environments to their children than those who were younger at baseline. For example, the average HOME-SF total score was 97.6 among mothers aged 20–22 at baseline compared with 90.6 among mothers aged 16 and 17. The same pattern held for the Emotional Support and Cognitive Stimulation subscales, though those aged 18 and 19 had the most favorable Harsh Discipline scores. The very young mothers in this sample appeared less able to provide a supportive and stimulating home environment.

Respondents who were age 16 or younger when they first gave birth reported providing *greater* warmth to the focal child (mean scores of 24.4), according to the Maternal Warmth/Responsiveness scale, than those who were older at the time of their first birth (22.9). However, women who gave birth at a comparatively young age received significantly lower scores on the Cognitive Stimulation subscale of the HOME-SF. It is interesting to note that the former is a subjective measure based on mothers' reports, while the latter is based on a combination of mothers' reports and interviewers' ratings.

Living Arrangements. Mothers who lived with either a husband or a partner at baseline scored less favorably on the HOME-SF total as well as on the Emotional Support and Physical Environment subscales than those who were neither married nor cohabiting at baseline. However, women who were in marital or nonmarital unions at baseline rated themselves as less "authoritarian" on the Maternal Control/Punitiveness scale than those with a partner or a husband (scores of 18.9 and 25.1, respectively). It is not clear why living with a husband or a partner would have these two rather contradictory associations with parenting measures among our sample of mothers. The presence of an additional adult in the home may provide the mother with support in the parenting role, but an unstable union may represent a source of conflict and distract the mother from her ability to parent effectively.

Education and Literacy. Higher levels of educational attainment were associated with higher Emotional Support, Cognitive Stimulation, Physical Environment, and HOME-SF total scores. Higher reading ability was associated with lower levels of restrictiveness on the Maternal Control/Punitiveness scale and generally more favorable Emotional Support scores on the HOME-SF. The positive association between educational attainment and literacy and favorable parenting is well documented in previous studies (for example, Risley and Hart, 1995).

Employment. Mothers who had no prior employment experience were rated lower on the HOME-SF total as well as the Physical Environment subscale. There is some evidence that certain qualities of the work environment, particularly the amount of self-direction, may influence the parenting practices of adults (see, for example, Kohn and Schooler, 1983). It may be that mothers who are homemakers exclusively have less exposure to a variety of enriching environments from which to draw upon as models to provide to their children.

Psychosocial Characteristics. Mothers with comparatively less favorable psychosocial well-being at baseline had less positive parenting practices. Women at high risk of depression had less favorable Emotional Support, Cognitive Stimulation, Harsh Discipline, and HOME-SF total scores on average than their counterparts. Mothers also scored less favorably on average on the Parenting Stress scale when at baseline they were more depressed. For example, the average score on the Parenting Stress scale among mothers who were at high risk of clinical depression at random assignment was 35.6 compared with 26.8 for those who were not at risk of depression.

Mothers with lower self-esteem had lower Emotional Support, Cognitive Stimulation, and HOME-SF total scores than did those with comparatively higher self-esteem scores. Negative self-esteem at baseline was also associated with higher levels of parenting stress.

Women with below-average Locus of Control scores provided less cognitively stimulating home environments to their children and scored less favorably on the HOME-SF total. Women with lower levels of internal Locus of Control also reported greater restrictiveness with their children and higher levels of parenting stress.

B. 18-Month Subgroup Scores on the Interview-Based Measures of Parenting

We may also ask whether the interview-based measures of parenting in this sample differed according to the characteristics of the mothers at the time of the 18-month follow-up interview. Tables 6.4–6.6 show results for average scores on the HOME-SF subscales, Maternal Report Parenting scales, and Time Use variables, respectively, for subgroups of women defined by their characteristics at the 18-month follow-up.

Child's Age. Mothers with the youngest children scored lowest on the Emotional Support subscale and the HOME-SF total. Since these analyses did not simultaneously control for mothers' age or other factors, it is unclear whether the subgroup differences in our sample are confounded with other attributes of mothers of young children. There may also be distinctive issues associated with completing the interview in families with younger as opposed to older children (see Part II of this monograph).

Living Arrangements. Women who were living with a husband or a partner at 18 months reported having lower levels of parenting stress and being less punitive with their children, and had higher Cognitive Stimulation subscale scores, but they also rated themselves less favorably on Maternal Warmth than did those who were neither married nor cohabiting at 18 months. As discussed previously, the presence of a partner in the home may alleviate some of the parental burden, but may also introduce additional sources of conflict and competition for the mother's time.

Education and Literacy. Mothers with the lowest literacy scores (7th grade and below) at the 18-month follow-up had significantly lower HOME-SF total, Emotional Support, and Cognitive Stimulation scores on average than mothers with higher literacy scores. Mothers with the lowest literacy scores also reported higher levels of parenting stress and more control and punitiveness in their parenting practices. Mothers with a high school diploma or trade school certificate, and even those who were in the process of obtaining a GED or high school diploma at the 18-month interview, had higher HOME-SF total scores than their counterparts with lower levels of educational attainment. They also scored more favorably on the Emotional Support, Cognitive Stimulation, and Physical Environment subscales. Comparatively higher educational attainment was also associated with lower Control/Punitiveness.

Table 6.4

Means on HOME-SF Scales, by 18-Month Follow-Up Characteristics

Subgroup Formed at 18 Months	n	Emotional Support	Cognitive Stimulation	Physical Environment	Harsh Discipline	Total
Full Sample						
Mean		96.68	98.36	97.61	97.92	96.30
Standard deviation		16.08	15.20	16.70	15.40	16.03
Range		50-127	58-135	51-118	39-119	45-129
Demographic Characteristics						
Child's age at observational study (months)						
24-36	84	92.99 **	96.99	96.43	97.03	93.32 *
37-47	112	97.75	97.43	96.42	99.39	96.49
48-63	82	98.99	99.77	99.54	96.60	99.11
Residence						
Living with partner/husband	54	96.37	103.90 ***	96.00	96.35	98.05
Not living with partner/husband	224	96.75	96.56	97.66	98.22	95.88
Living with parent/grandparent	98	96.26	98.87	98.64	98.17	97.32
Not living with parent/grandparent	180	96.90	97.50	96.64	97.39	95.75
Education and Literacy						
TABE reading score (grade equivalent)[a]						
7th grade or below	134	94.82 *	93.73 ***	96.00	97.63	92.82 ***
8th or 9th	67	99.79 *	101.26 **	98.81	99.74	100.21 *
10th or above	75	96.90	102.36 ***	98.00	96.29	98.51
Mother's education at 18 months						
No high school diploma or GED, not in school	145	93.79 ***	95.21 ***	95.44 *	97.71	92.87 ***
In process of getting GED or high school diploma	28	96.21	97.50	94.64	96.90	94.91
Has GED	55	97.43	100.89	100.35	100.10	99.85
High school or trade school diploma/attending college	50	104.50	103.10	101.05	96.37	103.16

(continued)

BEST COPY AVAILABLE

Table 6.4 (continued)

Subgroup Formed at 18 Months	n	Emotional Support	Cognitive Stimulation	Physical Environment	Harsh Discipline	Total
Earnings from Employment						
Respondent's total earnings over the 18-month period[b]						
No earnings	161	95.42	97.38	96.35	99.03	95.30
Below mean	81	97.30	97.13	99.21	95.94	96.68
At or above mean ($3,323.87)	36	100.93	102.60	97.56	96.95	99.95
Total earnings from respondent and partner or others during month 18[b]						
No earnings	204	95.66	96.47 **	96.96	98.04	95.05 *
Below mean	40	100.37	100.32	96.82	98.07	98.93
At or above mean ($719.40)	34	98.45	104.32	100.21	96.50	100.72
Fertility						
Pregnancy during follow-up	159	94.87 **	96.84	94.65 ***	97.81	93.83 ***
No pregnancy during follow-up	119	99.10	99.51	100.94	97.92	99.61
Birth during follow-up	98	92.72 ***	95.24	94.32 **	97.11 **	91.77 ***
No birth during follow-up	180	98.83	99.50	98.99	98.26	98.77
Abortion during follow-up	41	98.49	96.65	95.55	100.61	96.20
No abortion during follow-up	237	96.37	98.22	97.65	97.38	96.32
Two or more pregnancies during follow-up	42	99.85	99.05	98.87	101.21	99.50
Did not have two or more pregnancies during follow-up	236	96.11	97.80	96.07	97.26	95.74
Sexually active at follow-up	234	96.86	97.89	97.14	97.91	96.30
Not sexually active at follow-up	42	95.58	99.05	98.24	98.39	96.68
Child Care[c]						
Any child care during follow-up	245	96.93	97.94	98.24	98.83 **	97.04
No child care during follow-up	25	95.01	97.76	93.51	91.66	92.19
Child ever in child care before age 1	80	97.60	99.79	97.22	99.03	97.68
Child not in care before age 1	190	96.39	97.13	98.05	97.81	96.13

(continued)

Table 6.4 (continued)

Subgroup Formed at 18 Months	n	Emotional Support	Cognitive Stimulation	Physical Environment	Harsh Discipline	Total
Number of months child was in day care/preschool						
Below mean	172	94.74 ***	96.41 **	95.60 **	98.31	94.02 ***
At or above mean (3.7 months)	106	99.82	100.54	100.17	97.13	100.02
Number of months child was in care by grandparent						
Below mean	196	95.91	97.67	96.35	97.24	95.35
At or above mean (2.9 months)	82	98.53	98.74	99.71	99.34	98.59
Psychosocial Characteristics						
CES-D (depression) score[d]						
0-15 (not at risk)	169	96.86 *	99.80 ***	97.94	99.36 *	97.81 ***
16-23 (at some risk)	63	99.37	98.73	98.20	96.76	97.98
24-60 (at high risk)	46	92.32	90.29	93.96	93.84	88.47
Mastery score[e]						
Below mean	115	97.43	94.31 ***	95.25 *	97.48	93.80 **
At or above mean (22)	163	96.15	100.58	98.82	98.13	98.07
Difficult Life Circumstances score[f]						
Below mean	136	97.16	99.01	97.92	98.76	97.40
At or above mean (2.8)	142	96.22	97.01	96.79	96.99	97.25
Number of social supports[g]						
0-2	197	96.03	97.84	96.91	96.60 **	95.43
3 or more	80	98.13	98.30	98.21	100.81	98.25
Contact with father of child						
Father babysat for child in past year	122	96.52	101.11 **	97.47	97.67	95.25
Father did not babysit for child in past year	155	96.83	95.41	97.40	97.99	97.65
Father took child overnight	135	96.11	99.17	97.23	98.02	96.57
Father did not take child overnight	142	97.24	96.73	97.62	97.69	96.06
Father saw child once or not at all	77	96.54	96.55	97.20	97.66	95.49
Father saw child more than once	200	96.75	98.44	97.52	97.92	96.62

(continued)

Table 6.4 (continued)

Subgroup Formed at 18 Months	n	Emotional Support	Cognitive Stimulation	Physical Environment	Harsh Discipline	Total
Substance Abuse						
Respondent ever high on alcohol in past month	102	97.60	97.48	97.29	97.36	96.40
Respondent never high on alcohol in past month	172	96.22	98.24	97.28	98.08	96.21
Respondent used drugs in past month[h]	42	96.46	96.09	94.95	98.12	94.10
Respondent did not use drugs in past month	230	96 80	98.31	97.83	97.75	96.73
Sample size	290					

Sources: New Chance 18-month follow-up survey.

Notes: Means are adjusted for experimental/control status in order to be descriptive of the overall sample.

Statistical significance levels are indicated as *** < = 1 percent, ** < = 5 percent, and * < = 10 percent, which refers to the probability that sample means are different from each other only because of random error.

For subgroups that have more than two levels, the placement of asterisks does not indicate where the significant difference lies, but only that there was a significant difference among the set of means (except for the TABE subgroup, where each level was compared to all other grade equivalences).

Calculations for this table used data for all 290 respondents for whom there were 18-month follow-up survey data, including those with values of zero for outcomes and New Chance enrollees (i.e., experimentals) who did not participate in the program. The sample size may fall slightly short of the number reported because of missing or unusable items from some respondents' questionnaires.

A modified version of the short form of the Home Observation for Measurement of the Environment (HOME) Inventory (first administered in the National Longitudinal Survey of Youth) was administered. Scores here were age-standardized to have a mean of 100 and a standard deviation of 15. The Harsh Discipline subscale was reverse-coded so that higher scores indicate less harsh discipline.

[a] The test used to measure reading ability was the reading part of the Tests of Adult Basic Education (TABE). Most sites administered the survey form of the test, but some administered the full reading test.

[b] Means were computed excluding values of zero.

[c] Regular child care was defined as an ongoing arrangement used while the mother was in school, in training, or working.

[d] The Center for Epidemiological Studies Depression (CES-D) Scale is a widely used measure of depression; scores can range from 0 to 60.

[e] The Mastery Scale measures sense of mastery over personal events; scores can range from 7 to 28.

[f] Scores represent total numbers of ongoing problems or stresses the respondent faces, of a list of 10 problems.

[g] Respondents were asked to indicate the total number of individuals to whom they could turn for support.

[h] Self-report measures that included use of marijuana, cocaine, crack, heroin, PCP, and ice.

Table 6.5

Means on Maternal Report Parenting Scales, by 18-Month Follow-Up Characteristics

Subgroup Formed at 18 Months	n	Maternal Warmth	Parenting Stress	Maternal Control/ Punitiveness
Full Sample				
Mean		23.60	29.61	23.94
Standard deviation		6.28	14.52	11.96
Range		0-30	0-68	0-60
Demographic Characteristics				
Child's age at observational study (months)				
24-36	84	23.08	29.70	22.37
37-47	114	24.11	29.98	24.12
48-63	89	23.60	28.74	25.44
Residence				
Living with partner/husband	54	22.24 *	26.19 *	19.89 ***
Not living with partner/husband	233	23.98	30.31	24.97
Living with parent/grandparent	100	23.68	30.43	24.65
Not living with parent/grandparent	187	23.63	29.04	23.66
Education and Literacy				
TABE reading score (grade equivalent)[a]				
7th grade or below	135	23.78	31.77 **	27.39 ***
8th or 9th	70	23.39	27.32	22.82
10th or above	80	23.68	27.76	19.38 ***
Mother's education at 18 months				
No high school diploma or GED, not in school	149	24.08	30.27	26.37 ***
In process of getting GED or high school diploma	28	22.90	28.97	23.27
Has GED	55	23.03	27.07	18.65
High school or trade school diploma/attending college	55	23.47	30.19	23.13
Earnings from Employment				
Respondent's total earnings over the 18-month period[b]				
No earnings	164	23.63	28.53	24.02
Below mean	86	23.85	31.54	25.08
At or above mean ($3,323.87)	37	23.28	29.27	21.36
Total earnings from respondent and partner or others during month 18[b]				
No earnings	209	23.85	28.53	24.70
Below mean	43	22.87	32.56	23.41
At or above mean ($719.40)	35	23.42	27.02	20.52

(continued)

Table 6.5 (continued)

Subgroup Formed at 18 Months	n	Maternal Warmth	Parenting Stress	Maternal Control/ Punitiveness
Fertility				
Pregnancy during follow-up	162	23.42	28.61	25.10 *
No pregnancy during follow-up	125	23.96	30.74	22.52
Birth during follow-up	102	22.92	28.20	23.51
No birth during follow-up	185	24.06	30.26	24.28
Abortion during follow-up	41	23.70	30.77	25.90
No abortion during follow-up	246	23.64	29.31	23.67
Two or more pregnancies during follow-up	41	23.38	26.70	26.79
Did not have two or more pregnancies during follow-up	246	23.70	30.00	23.53
Sexually active at follow-up	241	23.67	29.72	24.29
Not sexually active at follow-up	44	23.36	27.91	21.27
Child Care[c]				
Any child care during follow-up	251	23.48	28.89 **	23.41 ***
No child care during follow-up	27	24.50	34.81	30.33
Child ever in child care before age 1	85	23.38	28.66	20.95
Child not in care before age 1	193	23.67	29.82	25.45
Number of months child was in day care/preschool				
Below mean	181	24.16 *	29.07	23.86
At or above mean (3.7 months)	106	22.76	30.32	24.25
Number of months child was in care by grandparent				
Below mean	201	23.57	29.68	24.18
At or above mean (2.9 months)	86	23.85	29.16	23.58
Psychosocial Characteristics				
CES-D (depression) score[d]				
0-15 (not at risk)	177	23.36	26.03 ***	22.80 *
16-23 (at some risk)	65	24.33	33.79	25.10
24-60 (at high risk)	45	23.81	37.14	27.15
Mastery score[e]				
Below mean	115	23.03	34.11 ***	26.50 ***
At or above mean (22)	172	24.07	26.47	22.34
Difficult Life Circumstances score[f]				
Below mean	142	23.89	26.48 ***	22.03 ***
At or above mean (2.8)	145	23.42	32.43	25.89
Number of social supports[g]				
0-2	200	23.44	29.24	23.93
3 or more	86	24.10	30.10	23.86

(continued)

Table 6.5 (continued)

Subgroup Formed at 18 Months	n	Maternal Warmth	Parenting Stress	Maternal Control/ Punitiveness
Contact with father of child				
Father babysat for child in past year	127	24.10	29.08	22.50 *
Father did not babysit for child in past year	159	23.30	29.97	25.27
Father took child overnight	147	23.65	30.92	23.34
Father did not take child overnight	139	23.63	28.35	24.76
Father saw child once or not at all	79	22.96	29.44	24.76
Father saw child more than once	207	23.91	29.64	23.81
Substance Abuse				
Respondent ever high on alcohol in past month	104	23.87	32.71 ***	26.18 **
Respondent never high on alcohol in past month	178	23.42	27.61	22.51
Respondent used drugs in past month[h]	43	24.22	35.13 ***	27.20 **
Respondent did not use drugs in past month	239	23.46	28.41	23.26
Sample size	290			

Sources: New Chance 18-month follow-up survey.

Notes: Means are adjusted for experimental/control status in order to be descriptive of the overall sample.

Statistical significance levels are indicated as *** < = 1 percent, ** < = 5 percent, and * < = 10 percent, which refers to the probability that sample means are different from each other only because of random error.

For subgroups that have more than two levels, the placement of asterisks does not indicate where the significant difference lies, but only that there was a significant difference among the set of means.

Calculations for this table used data for all 290 respondents for whom there were 18-month follow-up survey data, including those with values of zero for outcomes and New Chance enrollees (i.e., experimentals) who did not participate in the program. The sample size may fall slightly short of the number reported because of missing or unusable items from some respondents' questionnaires.

The three self-report parenting scales were developed for this study.

[a] The test used to measure reading ability was the reading part of the Tests of Adult Basic Education (TABE). Most sites administered the survey form of the test, but some administered the full reading test.

[b] Means were computed excluding values of zero.

[c] Regular child care was defined as an ongoing arrangement used while the mother was in school, in training, or working.

[d] The Center for Epidemiological Studies Depression (CES-D) Scale is a widely used measure of depression; scores can range from zero to 60.

[e] The Mastery Scale measures sense of mastery over personal events; scores can range from 7 to 28.

[f] Scores represent total numbers of ongoing problems or stresses the respondent faces, of a list of 10 problems.

[g] Respondents were asked to indicate the total number of individuals to whom they could turn for support.

[h] Self-report measures that included use of marijuana, cocaine, crack, heroin, PCP, and ice.

Table 6.6

Means on Parenting Time Use Measures, by 18-Month Follow-Up Characteristics

Subgroup Formed at 18 Months	n	Overall Parenting Time	Parenting Chore Time
Full Sample			
Mean		2.25	1.15
Standard deviation		2.23	1.23
Range		0-11	0-9
Demographic Characteristics			
Child's age at observational study (months)			
24-36	86	1.92	1.01
37-47	115	2.36	1.16
48-63	89	2.41	1.26
Residence			
Living with partner/husband	55	2.49	1.26
Not living with partner/husband	235	2.19	1.12
Living with parent/grandparent	101	2.21	1.09
Not living with parent/grandparent	189	2.26	1.18
Education and Literacy			
TABE reading score (grade equivalent)[a]			
7th grade or below	138	2.11	1.10
8th or 9th	70	2.39	1.33
10th or above	80	2.27	1.07
Mother's education at 18 months			
No high school diploma or GED, not in school	151	2.24	1.17
In process of getting GED or high school diploma	29	1.83	1.07
Has GED	55	2.48	1.12
High school or trade school diploma/attending college	55	2.25	1.16
Earnings from Employment			
Respondent's total earnings over the 18-month period[b]			
No earnings	167	2.17	1.17
Below mean	86	2.40	1.07
At or above mean ($3,323.87)	37	2.23	1.16
Total earnings from respondent and partner or others during month 18[b]			
No earnings	212	2.20	1.17
Below mean	43	2.30	1.05
At or above mean ($719.40)	35	2.43	1.13
Fertility			
Pregnancy during follow-up	165	2.50 **	1.35 ***
No pregnancy during follow-up	125	1.91	0.88

(continued)

Table 6.6 (continued)

Subgroup Formed at 18 Months	n	Overall Parenting Time	Parenting Chore Time
Birth during follow-up	103	2.57 *	1.35 **
No birth during follow-up	187	2.06	1.04
Abortion during follow-up	42	2.15	1.06
No abortion during follow-up	248	2.26	1.16
Two or more pregnancies during follow-up	42	2.62	1.57 **
Did not have two or more pregnancies during follow-up	248	2.18	1.08
Sexually active at follow-up	44	2.18	1.14
Not sexually active at follow-up	244	2.60	1.20
Child Care[c]			
Any child care during follow-up	253	2.13	1.10
No child care during follow-up	27	2.75	1.17
Child ever in child care before age 1	86	1.97	1.13
Child not in care before age 1	194	2.28	1.10
Number of months child was in day care/preschool			
Below mean	182	2.29	1.14
At or above mean (3.7 months)	108	2.17	1.17
Number of months child was in care by grandparent			
Below mean	203	2.41 *	1.20
At or above mean (2.9 months)	87	1.86	1.02
Psychosocial Characteristics			
CES-D (depression) score[d]			
0-15 (not at risk)	177	2.33	1.16
16-23 (at some risk)	67	2.04	1.01
24-60 (at high risk)	46	2.23	1.31
Mastery score[e]			
Below mean	117	2.32	1.28
At or above mean (22)	173	2.19	1.06
Difficult Life Circumstances score[f]			
Below mean	143	2.21	1.06
At or above mean (2.8)	147	2.28	1.23
Number of social supports[g]			
0-2	202	2.31	1.13
3 or more	87	2.09	1.19

(continued)

BEST COPY AVAILABLE

Table 6.6 (continued)

Subgroup Formed at 18 Months	n	Overall Parenting Time	Parenting Chore Time
Contact with father of child			
Father babysat for child in past year	127	2.46	1.28
Father did not babysit for child in past year	162	2.09	1.05
Father took child overnight	139	2.33	1.25
Father did not take child overnight	150	2.18	1.06
Father saw child once or not at all	82	2.24	1.17
Father saw child more than once	207	2.25	1.14
Substance Abuse			
Respondent ever high on alcohol in past month	106	2.05	1.05
Respondent never high on alcohol in past month	179	2.38	1.24
Respondent used drugs in past month[h]	43	2.76	1.24
Respondent did not use drugs in past month	242	2.16	1.15
Sample size	290		

Sources: New Chance 18-month follow-up survey and brief interview accompanying the observational session.

Notes: Means are adjusted for experimental/control status in order to be descriptive of the overall sample.

Statistical significance levels are indicated as *** < = 1 percent, ** < = 5 percent, and * < = 10 percent, which refers to the probability that sample means are different from each other only because of random error.

For subgroups that have more than two levels, the placement of asterisks does not indicate where the significant difference lies, but only that there was a significant difference among the set of means.

Calculations for this table used data for all 290 respondents for whom there were data from the brief interview accompanying the observational session and 18-month follow-up survey data, including those with values of zero for outcomes and New Chance enrollees (i.e., experimentals) who did not participate in the program. The sample size may fall slightly short of the number reported because of missing or unusable items from some respondents' questionnaires.

[a] The test used to measure reading ability was the reading part of the Tests of Adult Basic Education (TABE). Most sites administered the survey form of the test, but some administered the full reading test.

[b] Means were computed excluding values of zero.

[c] Regular child care was defined as an ongoing arrangement used while the mother was in school, in training, or working.

[d] The Center for Epidemiological Studies Depression (CES-D) Scale is a widely used measure of depression; scores can range from zero to 60.

[e] The Mastery Scale measures sense of mastery over personal events; scores can range from 7 to 28.

[f] Scores represent total numbers of ongoing problems or stresses the respondent faces, of a list of 10 problems.

[g] Respondents were asked to indicate the total number of individuals to whom they could turn for support.

[h] Self-report measures that included use of marijuana, cocaine, crack, heroin, PCP, and ice.

Economic Variables. Respondents with no earnings of their own or from others at 18 months had lower Cognitive Stimulation and HOME-SF total scores.

Fertility. Women who had had no pregnancies or births during the follow-up period provided higher-quality home environments (Emotional Support, Physical Environment, and HOME-SF total) for their children than those who either became pregnant or gave birth during the follow-up. In addition, expectant mothers at 18 months reported more Control/Punitiveness toward the focal child than those who were not pregnant. However, there was a differential in *quantity* of parenting that favored those with additional fertility experiences. Women who were pregnant at the time of the first follow-up spent more time parenting overall and logged more time on parenting chores than those who were not pregnant at 18 months. Similarly, women who had had a birth spent slightly more time doing parenting-related activities overall, including parenting chores.

Child Care. Mothers who reported that the focal child received *any* child care during the follow-up reported less Parenting Stress and less Control/Punitiveness toward their children. Mothers of children who spent a greater number of months in day care or preschool during the follow-up period obtained higher HOME-SF total, Cognitive Stimulation, Emotional Support, and Physical Environment scores. However, when children had spent fewer than the average number of months in day care or preschool at the 18-month follow-up, their mothers perceived themselves as higher in Maternal Warmth. There was a fairly sizable difference in the measure of Overall Parenting Time between those mothers whose children spent relatively less time in the care of a grandparent and those who spent relatively more time in the care of a grandparent at the 18-month follow-up. However, significant differences in Overall Parenting Time did not emerge according to other child care experiences.

Psychosocial Characteristics. As was the case at baseline, women at greater risk of depression at the 18-month follow-up, and who were rated less favorably on other measures of psychosocial adjustment, had significantly poorer parenting practices. There was a statistical association between risk of depression and less favorable HOME-SF total scores and subscale scores except for Physical Environment. Depression was also linked with greater parenting stress and more controlling or punitive disciplinary practices.

Above-average mastery scores were associated with significantly higher HOME-SF scores (Cognitive Stimulation, Physical Environment, and total) and lower levels of Parenting Stress and Control/Punitiveness. By contrast, above-average scores on the Difficult Life Circumstances measure were associated with more Parenting Stress and Control/Punitiveness.

Mothers who reported being high on alcohol or using drugs in the past months perceived themselves as under greater parenting stress and as being more controlling or punitive toward their children.

Father Contact. The involvement of fathers in babysitting in the past year was associated with higher Cognitive Stimulation scores and lower Control/Punitiveness scores. A cooperative relationship with the child's father may not only serve as a resource to the mother, but may also indicate positive attributes about the mother that she also brings to her parenting role.

III. Impacts of New Chance on the Interview-Based Measures of Parenting

A. Overall Group Differences

Table 6.7 presents impacts of the New Chance Program on each of the interview-based measures of parenting for the observational study sample. Statistically significant group differences were observed on a number of the interview-based parenting variables. Experimental group mothers in the observational study sample had higher HOME-SF total scores, on average, than those in the control group (a pattern not found in the full evaluation sample). Experimental group mothers were also rated more favorably on the Emotional Support subscale of the HOME-SF, a finding observed for the full New Chance sample at 18 months as well. There were no significant program impacts for the other HOME-SF subscales.

No program impacts were observed for either the Maternal Control/Punitiveness scale or the Parenting Stress scale in the observational study sample. However, New Chance experimentals had significantly higher scores on the Warmth/Responsiveness scale in this sample, rating themselves nearly two points higher on average than controls. This pattern of findings differs somewhat from that in the full evaluation sample at 18 months, in which a group difference was documented for the Control/Punitiveness scale rather than the Warmth scale.

Finally, mothers in the experimental group of the observational study sample spent more Overall Parenting Time and more Parenting Chore Time than mothers in the control group. Mothers in the experimental group spent approximately 48 minutes more in Overall Parenting Time than those in the control group and 24 minutes more specifically engaged in Parenting Chores. No parallel measure exists for the full evaluation sample.

B. Impacts for Baseline Subgroups

Our next question is that of whether the impacts of the New Chance Program detected on the interview-based measures of parenting occured across subgroups of mothers and their children or rather were restricted to specific baseline subgroups. We explored this issue by examining impacts on the interview-based measures of parenting in light of the demographic and socioeconomic characteristics of the families at baseline.

Tables 6.8–6.10 present impacts on the HOME Emotional Support subscale, Maternal Warmth scale, and Parenting Chore Time measure for subgroups defined by their baseline characteristics. The experimental-control differential on each of these measures of parenting is shown in the column labeled "within-subgroup impact," while the probability that the observed difference between experimentals and controls within the subgroup in question could have arisen by chance or simply reflects random error of measurement appears in the adjacent column labeled "p." We omit the HOME total score and the Overall Parenting Time measure in Tables 6.8–6.10 because these are viewed as summary scores; findings for the more specific scales related to each of these summary scores (the Emotional Support subscale of the HOME-SF and the Parenting Chore Time measure) are considered more informative.

Table 6.8 reveals a striking result: mean scores for experimentals on the Emotional Support subscale were numerically higher than controls in *all* the subgroups examined. Statistically significant within-group impacts were observed across the board for almost every subgroup cate-

Table 6.7

Impacts of New Chance on Survey Measures of Parenting at 18-Month Follow-Up

Measure	Experimentals	Controls	Difference	p[a]
HOME-SF				
Cognitive Stimulation	98.81	97.41	1.50	0.415
Emotional Support	99.54	91.58	7.96 ***	0.000
Physical Environment	98.29	96.42	1.87	0.373
Harsh Discipline	97.64	98.43	0.79	0.689
Total HOME Scale	98.00	93.28	4.72 **	0.017
Maternal Report				
Maternal Control/Punitiveness	23.56	24.59	-1.03	0.462
Maternal Warmth	24.21	22.57	1.64 **	0.038
Parenting Stress	30.19	28.89	1.30	0.538
Parenting Time Use				
Parenting Chore Time	1.29	0.89	0.40 **	0.013
Overall Parenting Time	2.54	1.74	0.80 ***	0.001
Sample size	184	106		

Source: New Chance 18-month follow-up survey and brief interview accompanying the observational session.

Notes: Calculations for this table used data for all 290 respondents for whom there were 18-month follow-up data, including those with values of zero for outcomes and New Chance enrollees (i.e., experimentals) who did not participate in the program. The sample sizes may fall slightly short of the numbers reported because of missing or unusable items from some respondents' questionnaires.

The averages are adjusted using linear analysis of covariance procedures controlling for seven kinds of difference in characteristics before random assignment (age of child, maternal literacy, whether mother had more than one child, gender of child, race/ethnicity, Philadelphia site, and Portland site).

[a]An F-test was applied to each difference in regression-adjusted means. The column labeled "p" is the statistical significance level of each between-group impact. That is, p is the probability that the averages for the experimental and control groups are different from each other only because of random error. Statistical significance levels are indicated as ***<=1 percent; **<=5 percent; *<= 10 percent.

Table 6.8

Impacts of New Chance on the HOME-SF Emotional Support Subscale, by Subgroups Formed at Baseline

Characteristics and Subgroup at Baseline	n	HOME-SF Emotional Support Experimentals	Controls	Within-Subgroup Impact	p[a]	Between-Subgroups Difference[b]	p[c]
Full Sample							
Mean		99.53	91.60	7.93 ***	0.000		
Demographic Characteristics							
Child's age at observational study (months)						--- *	0.096
24-36	84	98.37	88.54	9.83 ***	0.005		
37-47	112	97.92	94.83	3.09	0.299		
48-63	82	103.09	90.46	12.63 ***	0.000		
Gender of child[d]						4.97	0.187
Female	139	100.85	90.52	10.33 ***	0.000		
Male	139	98.30	92.94	5.36 *	0.051		
Number of children						-2.66	0.498
1	177	99.42	92.49	6.93 ***	0.005		
2 or more	101	99.87	90.28	9.59 ***	0.002		
Race/ethnicity						-2.64	0.660
Black	233	99.49	91.82	7.67 ***	0.000		
White	45	99.69	89.38	10.31 *	0.069		
Age of mother (years)						---	0.144
16-17	54	94.69	92.64	2.05	0.632		
18-19	141	102.83	91.47	10.36 ***	0.000		
20-22	83	97.04	90.79	6.25 *	0.067		
Living arrangement (mother)						3.46	0.368
Living with mother	103	102.52	92.49	10.03 ***	0.001		
Not living with mother	168	92.12	91.55	6.57 ***	0.008		
Living arrangement (partner/husband)						3.78	0.564
Living with partner/husband	29	95.68	84.34	11.34 *	0.072		
Not living with partner/husband	242	100.21	92.65	7.56 ***	0.000		
Age at first child's birth (years)						3.92	0.306
13-16	129	101.52	91.51	10.01 ***	0.000		
17-19	149	97.89	91.80	6.09 **	0.022		
Age of youngest child (years)						4.07	0.297
Under 1	103	100.87	90.45	10.42 ***	0.001		
1 or older	175	98.84	92.49	6.35 ***	0.010		

(continued)

Table 6.8 (continued)

Characteristics and Subgroup at Baseline	n	HOME-SF Emotional Support Experimentals	 Controls	Within-Subgroup Impact	p[a]	Between-Subgroups Difference[b]	p[c]
Number of pregnancies						-1.57	0.717
1 or 2	208	98.59	91.27	7.32 ***	0.001		
3 or more	69	102.48	93.59	8.89 **	0.020		
Education and Literacy							
Highest grade completed						-3.71	0.338
10th or below	178	97.58	90.78	6.80 ***	0.005		
11th or above	100	103.36	92.85	10.51 ***	0.001		
Interval since last attended regular high school						-3.70	0.324
2 years or less	131	99.36	93.05	6.31 **	0.021		
3 years or more	140	100.08	90.07	10.01 ***	0.000		
TABE reading test score (grade equivalent)[e]						---	0.347
7th grade or below	134	96.62	90.64	5.98 **	0.036		
8th or 9th grade	67	102.31	94.47	7.84 **	0.033		
10th grade or above	75	102.56	90.07	12.49 ***	0.000		
Employment and AFDC History							
Ever employed						-3.49	0.439
Yes	214	100.26	93.14	7.12 ***	0.008		
No	64	97.09	86.48	10.61 ***	0.001		
Family received AFDC when sample member was growing up						---	0.369
Never	79	100.15	91.71	8.84 **	0.029		
Sometimes	138	100.51	90.75	9.76 ***	0.000		
Always	60	96.21	93.16	3.05	0.442		
Psychosocial Characteristics							
CES-D (depression) score[f]						---	0.610
0-15 (not at risk)	134	99.50	91.64	8.06 ***	0.004		
16-23 (at some risk)	74	100.36	94.74	5.62	0.125		
24-60 (at high risk)	70	98.73	87.93	10.80 ***	0.005		
Self-Esteem score[g]						8.32 *	0.056
Below mean	73	99.62	85.37	14.25 ***	0.000		
At or above mean (35)	205	99.50	93.57	5.93 ***	0.007		
Number of sources of emotional support						0.27	0.944
0-2	125	99.92	91.78	8.14 ***	0.004		
3 or more	153	99.26	91.39	7.87 ***	0.003		

(continued)

Table 6.8 (continued)

Characteristics and Subgroup at Baseline	n	HOME-SF Emotional Support		Within-Subgroup Impact	p[a]	Between-Subgroups Difference[b]	p[c]
		Experimentals	Controls				
Level of satisfaction with emotional support[h]						-3.41	0.365
Less than very satisfied	131	99.39	93.18	6.21 **	0.025		
Very satisfied	147	99.67	90.05	9.62 ***	0.000		
Locus of Control score[i]						-2.95	0.444
Below mean	112	99.29	93.07	6.22 **	0.036		
At or above mean (22)	166	99.73	90.56	9.17 ***	0.000		
Sample size	290						

Sources: New Chance baseline enrollment data and 18-month follow-up survey.

Notes: Calculations for this table used data for all 290 respondents for whom there were 18- month follow-up survey data, including those with values of zero for outcomes and New Chance enrollees (i.e., experimentals) who did not participate in the program. The sample size may fall slightly short of this number because of missing or unusable items from some respondents' questionnaires.

The averages are adjusted using linear analysis of covariance procedures controlling for up to seven kinds of difference in characteristics before random assignment (age of child, maternal literacy, whether mother had more than one child, gender of child, race/ethnicity, Philadelphia site, and Portland site).

[a] A two-tailed t-test was applied to each difference in regression-adjusted means. The column labeled "p" is the statistical significance level of each within-subgroup impact. That is, p is the probability that sample estimates are different from zero or from each other only because of random error. Statistical significance levels are indicated as ***<=1 percent; **<=5 percent and ; *<= 10 percent.

[b] For each characteristic with only two subgroups, the between-subgroups difference is the impact for the first subgroup less the impact for the second subgroup. For characteristics with more than two subgroups, no single difference between subgroup impacts can be calculated, as indicated by dashes in the table. However, it is possible to assess the statistical significance of variation across multiple subgroups, as indicated by the asterisks.

[c] An F-test was applied to the interaction between the subgroup and experimental/control status. The column labeled "p" is the statistical significance level of each between-subgroups difference in impacts. That is, p is the probability that sample estimates are different from zero or from each other only because of random error. Statistical significance levels are indicated as ***<=1 percent; **<=5 percent; and *<= 10 percent.

[d] When a sample member had more than one child, her response refers to the randomly selected focal child.

[e] The test used to measure reading ability was the reading part of the Tests of Adult Basic Education (TABE). Most sites administered the survey form of the test, but some administered the full reading test.

[f] The Center for Epidemiological Studies Depression (CES-D) Scale is a widely used measure of depression; scores can range from zero to 60.

[g] The measure of self-esteem used was the Rosenberg Self-Esteem Scale, a 10-item scale that assesses a person's global sense of self-worth. Scores can range from 10 to 50; 30 is considered the neutral midpoint.

[h] Enrollees were asked about their degree of satisfaction with the emotional support ("people who listen to you, reassure you, and show you they care") they received. Levels range from 1 (very dissatisfied) to 5 (very satisfied); however, because so few individuals indicated that they were less than very satisfied with their emotional support, levels 1 through 4 were collapsed into one category labeled "less than very satisfied."

[i] The Locus of Control Scale is a six-item adaptation of the longer scale originally developed by Julien Rotter (1966). Scores can range from 6 to 30; 18 is considered the neutral midpoint.

Table 6.9

Impacts of New Chance on the Maternal Warmth Scale, by Subgroups Formed at Baseline

Characteristics and Subgroup at Baseline	n	Maternal Warmth		Within-Subgroup Impact		Between-Subgroups Difference	
		Experimentals	Controls		p^a		p^b
Full Sample							
Mean		24.17	22.63	1.54 **	0.048		
Demographic Characteristics							
Child's age at observational study (months)						---	0.416
24-36	84	24.06	21.84	2.22	0.126		
37-47	114	24.16	23.81	0.35	0.779		
48-63	89	24.43	21.76	2.67 **	0.058		
Gender of child[c]						-1.47	0.345
Female	141	24.10	23.21	0.89	0.428		
Male	146	24.34	21.98	2.36 **	0.032		
Number of children						1.25	0.440
1	181	24.00	21.87	2.13 **	0.036		
2 or more	106	24.52	23.67	0.88	0.483		
Race/ethnicity						5.42 **	0.010
Black	242	24.70	23.37	1.33 ***	0.005		
White	45	21.92	26.01	-4.09 *	0.083		
Age of mother (years)						---	0.678
16-17	55	25.75	22.82	2.93	0.103		
18-19	147	23.80	22.28	1.52	0.171		
20-22	85	23.84	22.95	0.89	0.533		
Living arrangement (mother)						1.99	0.212
Living with mother	103	24.44	21.67	2.77 **	0.027		
Not living with mother	177	24.17	23.39	0.78	0.433		
Living arrangement (partner/husband)						-1.37	0.607
Living with partner/husband	29	22.66	22.32	0.34	0.894		
Not living with partner/husband	251	24.46	22.75	1.71 **	0.037		
Age at first child's birth (years)						-1.22	0.437
13-16	133	24.39	23.36	1.03	0.361		
17-19	154	24.04	21.79	2.25 **	0.041		
Age of youngest child (years)						1.91	0.238
Under 1	106	24.17	21.40	2.77 **	0.028		
1 or older	181	24.25	23.39	0.86	0.393		

(continued)

Table 6.9 (continued)

Characteristics and Subgroup		Maternal Warmth		Within-Subgroup		Between-Subgroups	
at Baseline	n	Experimentals	Controls	Impact	p[a]	Difference	p[b]
Number of pregnancies						2.35	0.194
1 or 2	216	24.46	22.35	2.11 **	0.020		
3 or more	70	23.36	23.60	-0.24	0.879		
Education and Literacy							
Highest grade completed						-0.29	0.850
10th or below	182	24.02	22.48	1.54	0.126		
11th or above	105	24.55	22.72	1.83	0.149		
Interval since last attended regular high school						2.16	0.166
2 years or less	133	25.21	22.43	2.78 **	0.015		
3 years or more	148	23.47	22.85	0.62	0.565		
TABE reading test score (grade equivalent)[d]						---	0.252
7th grade or below	135	24.23	22.22	2.01 *	0.089		
8th or 9th grade	70	24.90	21.85	3.05 **	0.040		
10th grade or above	80	23.44	23.71	-0.27	0.854		
Employment and AFDC History							
Ever employed						-2.00	0.282
Yes	223	23.75	22.52	1.23 **	0.050		
No	64	25.91	55.68	3.23	0.173		
Family received AFDC when sample member was growing up						---	0.925
Never	83	23.40	21.28	2.12	0.167		
Sometimes	145	24.71	23.32	1.39	0.208		
Always	58	24.20	22.45	1.75	0.300		
Psychosocial Characteristics							
CES-D (depression) score[e]						---	0.152
0-15 (not at risk)	139	25.11	21.99	3.12 ***	0.006		
16-23 (at some risk)	77	23.02	22.13	0.89	0.551		
24-60 (at high risk)	71	23.77	24.22	-0.45	0.773		
Self-Esteem score[f]						-4.56 **	0.011
Below mean	74	22.97	24.76	-1.79	0.247		
At or above mean (35)	213	24.64	21.87	2.77 ***	0.002		
Number of sources of emotional support						-0.11	0.945
0-2	132	23.67	22.17	1.50	0.183		
3 or more	155	24.61	23.00	1.61	0.143		

(continued)

Table 6.9 (continued)

Characteristics and Subgroup at Baseline	n	Maternal Warmth: Experimentals	Maternal Warmth: Controls	Within-Subgroup Impact	p[a]	Between-Subgroups Difference	p[b]
Level of satisfaction with emotional support[g]						1.46 *	0.087
Less than very satisfied	137	23.88	21.55	2.33 **	0.039		
Very satisfied	150	24.48	23.61	0.87	0.418		
Locus of Control score[h]						-1.28	0.420
Below mean	114	23.95	23.08	0.87	0.476		
At or above mean (22)	173	24.38	22.23	2.15 **	0.035		
Sample size	290						

Sources: New Chance baseline enrollment data and 18-month follow-up survey.

Notes: Calculations for this table used data for all 290 respondents for whom there were 18-month follow-up survey data, including those with values of zero for outcomes and New Chance enrollees (i.e., experimentals) who did not participate in the program. The sample size may fall slightly short of this number because of missing or unusable items from some respondents' questionnaires.

The averages are adjusted using linear analysis of covariance procedures controlling for up to seven kinds of difference in characteristics before random assignment (age of child, maternal literacy, whether mother had more than one child, gender of child, race/ethnicity, Philadelphia site, and Portland site).

[a] A two-tailed t-test was applied to each difference in regression-adjusted means. The column labeled "p" is the statistical significance level of each within-subgroup impact. That is, p is the probability that sample estimates are different from zero or from each other only because of random error. Statistical significance levels are indicated as ***<=1 percent; **<=5 percent and ; *<= 10 percent.

[b] An F-test was applied to the interaction between the subgroup and experimental/control status. The column labeled "p" is the statistical significance level of each between-subgroups difference in impacts. That is, p is the probability that sample estimates are different from zero or from each other only because of random error. Statistical significance levels are indicated as ***<=1 percent; **<=5 percent; and *<= 10 percent.

[c] When a sample member had more than one child, her response refers to the randomly selected focal child.

[d] The test used to measure reading ability was the reading part of the Tests of Adult Basic Education (TABE). Most sites administered the survey form of the test, but some administered the full reading test.

[e] The Center for Epidemiological Studies Depression (CES-D) Scale is a widely used measure of depression; scores can range from zero to 60.

[f] The measure of self-esteem used was the Rosenberg Self-Esteem Scale, a 10-item scale that assesses a person's global sense of self-worth. Scores can range from 10 to 50; 30 is considered the neutral midpoint.

[g] Enrollees were asked about their degree of satisfaction with the emotional support ("people who listen to you, reassure you, and show you they care") they received. Levels range from 1 (very dissatisfied) to 5 (very satisfied); however, because so few individuals indicated that they were less than very satisfied with their emotional support, levels 1 through 4 were collapsed into one category labeled "less than very satisfied."

[h] The Locus of Control Scale is a six-item adaptation of the longer scale originally developed by Julien Rotter (1966). Scores can range from 6 to 30; 18 is considered the neutral midpoint.

Table 6.10

Impacts of New Chance on the Parenting Chore Time Measure, by Subgroups Formed at Baseline

Characteristics and Subgroup at Baseline	n	Parenting Chore Time		Within-Subgroup Impact	p[a]	Between-Subgroups Difference	p[b]
		Experimentals	Controls				
Full Sample							
Mean		1.29	0.89	0.40 ***	0.000		
Demographic Characteristics							
Child's age at observational study (months)							
24-36	86	1.19	0.82	0.37	0.209	---	0.974
37-47	115	1.29	0.93	0.36	0.147		
48-63	89	1.38	0.94	0.44	0.119		
Gender of child[c]						0.26	0.440
Female	142	1.33	0.81	0.53 **	0.020		
Male	148	1.26	0.99	0.27	0.226		
Number of children						-0.58 *	0.076
1	183	1.11	0.94	0.17	0.407		
2 or more	107	1.63	0.88	0.75 ***	0.003		
Race/ethnicity						-0.39	0.445
Black	244	1.27	0.91	0.36 **	0.036		
White	46	1.40	0.65	0.75	0.123		
Age of mother (years)						---	0.895
16-17	56	1.21	0.90	0.31	0.389		
18-19	148	1.24	0.76	0.48 **	0.034		
20-22	86	1.45	1.11	0.34	0.242		
Living arrangement (mother)						-0.17	0.586
Living with mother	104	1.31	1.08	0.23	0.343		
Not living with mother	179	1.18	0.78	0.40 **	0.040		
Living arrangement (partner/husband)						0.62	0.231
Living with partner/husband	30	1.20	0.31	0.89 *	0.073		
Not living with partner/husband	253	1.23	0.96	0.27 *	0.091		
Age at first child's birth (years)						0.20	0.509
13-16	135	1.28	0.78	0.50 **	0.028		
17-19	155	1.31	1.01	0.30	0.191		
Age of youngest child (years)						0.17	0.597
Under 1	107	1.15	0.65	0.50	0.052		
1 or older	183	1.38	1.05	0.33	0.116		

(continued)

Table 6.10 (continued)

Characteristics and Subgroup at Baseline	n	Parenting Chore Time		Within-Subgroup Impact	p[a]	Between-Subgroups Difference	p[b]
		Experimentals	Controls				
Number of pregnancies						-0.53	0.159
1 or 2	218	1.24	0.98	0.26	0.150		
3 or more	71	1.47	0.68	0.79 **	0.015		
Education and Literacy							
Highest grade completed						-0.25	0.453
10th or below	185	1.20	0.89	0.31	0.123		
11th or above	105	1.47	0.91	0.56 **	0.031		
Interval since last attended regular high school						0.38	0.232
2 years or less	134	1.38	0.79	0.59 **	0.012		
3 years or more	149	1.21	1.00	0.21	0.350		
TABE reading test score (grade equivalent)[d]						---	0.824
7th grade or below	132	1.30	0.83	0.47 **	0.051		
8th or 9th grade	79	1.31	1.07	0.24	0.428		
10th grade or above	79	1.26	0.82	0.44	0.136		
Employment and AFDC History							
Ever employed						0.24	0.536
Yes	224	1.35	0.90	0.45 **	0.014		
No	66	1.09	0.88	0.21	0.524		
Family received AFDC when sample member was growing up						---	0.495
Never	84	1.31	1.08	0.23	0.459		
Sometimes	145	1.38	0.81	0.57 **	0.012		
Always	60	1.05	0.89	0.16	0.652		
Psychosocial Characteristics							
CES-D (depression) score[e]						---	0.380
0-15 (not at risk)	141	1.22	0.87	0.35	0.129		
16-23 (at some risk)	77	1.13	0.98	0.15	0.622		
24-60 (at high risk)	72	1.60	0.85	0.75 **	0.019		
Self-Esteem score[f]						0.55	0.129
Below mean	215	1.63	0.82	0.81 ***	0.010		
At or above mean (35)	75	1.18	0.92	0.26	0.154		
Number of sources of emotional support						0.22	0.488
0-2	133	1.34	0.83	0.51 **	0.026		
3 or more	157	1.26	0.97	0.29	0.191		

(continued)

Table 6.10 (continued)

Characteristics and Subgroup at Baseline	n	Parenting Chore Time Experimentals	Parenting Chore Time Controls	Within-Subgroup Impact	p^a	Between-Subgroups Difference	p^b
Level of satisfaction with emotional support[g]						0.52	0.384
Less than very satisfied	138	1.51	0.83	0.68 **	0.003		
Very satisfied	152	1.11	0.95	0.16	0.462		
Locus of Control score[h]						0.12	0.695
Below mean	115	1.39	0.92	0.47 *	0.058		
At or above mean (22)	175	1.23	0.88	0.35 *	0.094		
Sample size	290						

Sources: New Chance baseline enrollment data, 18-month follow-up survey, and brief interview accompanying observational session.

Notes: Calculations for this table used data for all 290 respondents for whom there were 18-month follow-up survey data, including those with values of zero for outcomes and New Chance enrollees (i.e., experimentals) who did not participate in the program. The sample size may fall slightly short of this number because of missing or unusable items from some respondents' questionnaires.

The averages are adjusted using linear analysis of covariance procedures controlling for up to seven kinds of difference in characteristics before random assignment (age of child, maternal literacy, whether mother had more than one child, gender of child, race/ethnicity, Philadelphia site, and Portland site).

[a] A two-tailed t-test was applied to each difference in regression-adjusted means. The column labeled "p" is the statistical significance level of each within-subgroup impact. That is, p is the probability that sample estimates are different from zero or from each other only because of random error. Statistical significance levels are indicated as ***<=1 percent; **<=5 percent and ; *<= 10 percent.

[b] An F-test was applied to the interaction between the subgroup and experimental/control status. The column labeled "p" is the statistical significance level of each between-subgroups difference in impacts. That is, p is the probability that sample estimates are different from zero or from each other only because of random error. Statistical significance levels are indicated as ***<=1 percent; **<=5 percent; and *<= 10 percent.

[c] When a sample member had more than one child, her response refers to the randomly selected focal child.

[d] The test used to measure reading ability was the reading part of the Tests of Adult Basic Education (TABE). Most sites administered the survey form of the test, but some administered the full reading test.

[e] The Center for Epidemiological Studies Depression (CES-D) Scale is a widely used measure of depression; scores can range from zero to 60.

[f] The measure of self-esteem used was the Rosenberg Self-Esteem Scale, a 10-item scale that assesses a person's global sense of self-worth. Scores can range from 10 to 50; 30 is considered the neutral midpoint.

[g] Enrollees were asked about their degree of satisfaction with the emotional support ("people who listen to you, reassure you, and show you they care") they received. Levels range from 1 (very dissatisfied) to 5 (very satisfied); however, because so few individuals indicated that they were less than very satisfied with their emotional support, levels 1 through 4 were collapsed into one category labeled "less than very satisfied."

[h] The Locus of Control Scale is a six-item adaptation of the longer scale originally developed by Julien Rotter (1966). Scores can range from 6 to 30; 18 is considered the neutral midpoint.

gory. Thus, the tendency for mothers in the experimental group to provide greater emotional support to their children than mothers in the control group was nearly universal within the observational study sample. Few differential patterns by mother's demographic characteristics, education and literacy, employment, or psychosocial characteristics emerged despite earlier evidence that scores on the Emotional Support subscale varied for those differing on these dimensions.

Table 6.9 presents comparable data for the Maternal Warmth scale. Scores were typically numerically higher for experimentals than for controls across subgroups examined. Yet the pattern of significant differences for subgroups is difficult to summarize. Most of the subgroups for which significant impacts occurred could be thought of as comparatively more advantageous characteristics at baseline — for example, having only one child, being older at the birth of the first child, having employment experience, being at low risk of depression, and having above-average Locus of Control. Yet other subgroups for which a significant impact was found point to greater disadvantage among mothers, such as lower TABE reading scores, and low satisfaction with emotional support.

Table 6.10 indicates that Parenting Chore Time was, again, numerically higher for mothers in the experimental group than among the controls for all of the subgroups examined. However, as for the Maternal Warmth scale, it is not easy to characterize the pattern of statistically significant differences for subgroups. Most of the instances of significant differences for subgroups are observed for variables indicative of greater disadvantage at baseline (for example, at high risk of depression, younger at the birth of the first child, and having below-average reading ability and lower self esteem), yet at the same time, a significant difference occurred within one subgroup indicating more favorable circumstances (11th or above for highest grade completed).

IV. Summary

Interview-based measures of parenting for the mothers in the observational study sample point to differences in both the quantity and quality of parenting. Experimental group mothers report spending more time on parenting activities overall, and on parenting chores in particular, such as bathing and feeding their children. The home environment was rated as more supportive and stimulating overall for mothers in the experimental group, and a difference was found particularly in the area of the emotional support available to the focal child in the home. The findings on the HOME-SF are in accord with the self-perceptions of the mothers: experimental group mothers perceived themselves as showing greater warmth to their children.

Given the number of individual scales or measures considered (five HOME, three Maternal Report Parenting, and two Time Use), the differences on the interview-based measures of parenting are fairly pervasive and exceed the number of differences that might be found by chance. Differences on the HOME Emotional Support subscale occurred for nearly every baseline subgroup. While scores were numerically higher on the Maternal Warmth and Parenting Chore Time measures for the experimental group families in virtually every subgroup, significant group differences were not systematically concentrated among those baseline subgroups pointing to either greater or less disadvantage at baseline.

It is interesting to note the absence of a significant difference on the interview-based measure of cognitive stimulation in the home (the HOME-SF Cognitive Stimulation subscale), in light of the observational finding of higher Book Reading Quality scores among New Chance experimental group mothers (see Chapter 5). The interview-based measure of cognitive stimulation focuses much more on the *amount* of cognitive stimulation the child receives than on the quality of that stimulation — for example, documenting the number of books in the house and the number of times per week the mother reads to the child. The possibility exists that participation in New Chance tended to enhance the quality, but not the quantity, of literacy-related interactions.

Finally, we note the basic concurrence across the observational and interview-based measures of parenting with regard to the socioemotional aspects of the parent-child relationship: results jointly suggest that the affective quality of interactions was somewhat more positive in families participating in New Chance. Yet we note a subtle difference between the observational and interview-based measures in this domain as well. While the direct observations of mother-child interaction showed a diminution in the occurrence of harsh interactions between mother and child for experimental group families, and no overall group difference in the supportiveness of interactions, the reverse pattern was found with the interview-based measures of parenting, with greater supportiveness but no indication of fewer harsh interactions noted for the experimental group. In light of the differing contexts of mother-child interaction that the observational and interview-based measures sample, this should not necessarily be seen as surprising or as a substantive disagreement. A challenging task, by increasing the frequency of requests or demands by the mother to the child, may distill out group differences in harsh treatment. By contrast, interactions in the course of everyday life, as documented through the HOME-SF and Maternal Report scales, may be more revealing of variation in the mothers' supportiveness.

Chapter 7

Integration: Looking Across the Differing Measures of Parenting

Martha J. Zaslow, M. Robin Dion, and Donna R. Morrison

This chapter looks across the differing measures of parenting available for the observational study sample, examining the consistency and magnitude of program impact findings, exploring the relationship between parenting behavior and program participation for mothers in the experimental group, and considering the extent to which the differing parenting measures overlap and provide distinct information. We note that there is at least one significant positive program impact for each of the five types of parenting measures included in the study: observational measures of the affective quality of interaction, observational measures of literacy interactions, Maternal Report Parenting scales, the abbreviated HOME Inventory, and Time Use measures. The fact that positive program impacts occurred across such different types of measures increases our confidence in the conclusion that New Chance did indeed affect parenting behavior. Yet we also note that impacts were found for only a minority of the parenting measures overall, and that the magnitude of impacts fell in the small-to-medium range of effect sizes. Parenting behavior within the experimental group was most closely linked with mothers' participation in the human capital development activities of the New Chance Program and in the program overall. There was also some indication of an association between parenting behavior and mothers' participation in parenting classes as well as in other program components with content related to parenting behavior.

While researchers have often portrayed observational measures as superior to interview-based measures of parenting, we argue here that parenting measures obtained in these two ways provide different and complementary information. The observational and interview measures show low-to-moderate intercorrelations, a finding that supports the view that these measures are related but not overlapping. Finally, analyses indicate that the observational measures of parenting add significantly to our ability to predict child outcomes beyond interview measures of parenting for two of five child outcomes examined.

Up to this point, separate consideration has been given to different sets of parenting measures. Chapter 4 focused on observational measures of the affective quality of mother-child interaction; Chapter 5 on observational measures related to the emergence of literacy, and Chapter 6 on measures of the mother-child relationship derived from interviews (both maternal report and interviewer ratings).

Yet some central questions require joint consideration of the different parenting measures. It is important to ask, for example, whether the New Chance Program had stronger program impacts for some aspects of parenting behavior than for others. It is important to explore the linkages between parenting behavior and participation in specific components of the New Chance Program, asking whether participation in the same program activities was associated with each of the documented program impacts on parenting. We need to ask how the interview-based and observational measures of parenting function as predictors of child outcomes and, further, whether the observational measures add to our ability to predict child outcomes beyond what would be possible if we were restricted to interview-based measures of parenting. These questions require that we examine all of the available parenting measures together rather than separately.

This chapter seeks to augment the presentation of findings to this point in the following ways:

- We focus first on what further can be learned about *program impacts* on parenting behavior through joint consideration of the parenting measures. We begin by summarizing the program impact findings, assessing the consistency of findings across parenting measures with very different characteristics. We then turn to new analyses exploring the magnitude of program impacts on parenting behavior, asking whether stronger effects were documented for some parenting measures than for others. New analyses also explore the extent to which positive program impacts on the differing parenting measures can be linked to mothers' participation in specific New Chance Program components.

- We then examine the extent to which the differing measures of parenting included in the New Chance Observational Study provide *distinct perspectives on the mother-child relationship* and how these measures of parenting *contribute to our ability to predict child outcomes at 42 months*. We begin with a conceptual assessment of the measures of parenting, examining how the measures differ (and are similar) in terms of the informant they rely on, the scope or breadth of the measures, and aspects of parenting addressed. We continue with an empirical assessment, examining correlations among the measures in order to address specific questions raised in the conceptual assessment. The chapter concludes with an analysis of interview-based and observational measures of parenting as predictors of selected child outcomes from the 42-month follow-up. We ask whether both interview-based and observational measures of parenting function as significant predictors of child outcomes. Further, we ask whether the observational measures add to our ability to predict child outcomes above and beyond what is possible when only interview-based measures of parenting are taken into account.

I. Consistency and Magnitude of Program Impacts on Parenting and Linkages with Program Participation

A. Consistency of Findings of Program Impacts on Parenting

The New Chance Observational Study encompassed (1) observational measures of the affective quality of mother-child interaction; (2) observational measures of the literacy-related aspects of mother-child interaction; (3) Maternal Report Parenting scales providing mothers' perspectives on the experience of parenting, focusing on warmth, stress, and control in the mother-child relationship; (4) an abbreviated version of the HOME Inventory (HOME-SF); and (5) measures of Time Use allocated to parenting (Overall Parenting Time and Parenting Chore Time).

It is important to note that previous chapters revealed at least one statistically significant positive program impact for each of these types of parenting measures. The observational measures of affective quality detected a difference in Harsh Treatment: mothers in the New Chance experimental group were less harsh in their interactions with their young children (Chapter 4). Among the observational measures of literacy interactions, mothers in the experimental group were rated higher on Book Reading Quality (Chapter 5). On the Maternal Report Parenting scales included among the interview-based measures of parenting, mothers in the experimental group rated themselves higher on the Warmth scale. HOME-SF findings indicated that the home environments provided by mothers in the experimental group were more supportive and stimulating overall (a difference on the HOME total score), and differed especially with regard to the provision of emotional support (a difference on the Emotional Support subscale). On the Time Use measures, mothers in the experimental group reported spending more time on parenting chores (Parenting Chore Time) and more time on parenting activities in general (Overall Parenting Time) (Chapter 6).

The fact that positive program impacts were found across such different parenting measures strengthens our confidence in the conclusion that New Chance affected parenting behavior positively. Positive impacts have been documented whether the source of data was a coder unaware of which research group a family was in (the observational measures), the mother alone (the Maternal Report Parenting scales and Time Use measures), or the mother in combination with a survey interviewer (the HOME-SF). Even though both sets of observational measures relied on coders, positive program impacts were found for observational measures coded in two independent research laboratories, following very different coding protocols. Further, findings pertain to the amount of time mothers spent with their children as well as to the quality of interactions when mothers and children were together. Group differences favoring the experimental group have been found for both cognitive and socioemotional aspects of maternal behavior.

While positive program impacts thus appear across different measurement approaches, a conclusion that positive program impacts appear quite consistently must be qualified in several ways. First, we must note that while there was at least one positive program impact for each type of parenting measure, the number of significant differences was nevertheless not great in light of the number of *discrete* measures within each *set* of measures. For example, of the five ratings of the affective quality of mother-child interaction selected for detailed examination, a significant

BEST COPY AVAILABLE

positive impact was found for only one. Similarly, of eight discrete measures of the literacy-related aspects of mother-child interaction, a positive program impact was found for only one. Thus, while the findings of positive program impacts can be considered consistent when looking across different types of parenting measures, they cannot be considered pervasive given the total number of discrete measures considered.

A second qualification comes from our analyses of whether the program impacts on parenting behavior occurred consistently across families with differing characteristics or were found only for definable subgroups. Only the program impact on the HOME Emotional Support subscale could be said to occur virtually universally. For the other parenting variables, significant group differences occurred for specific baseline subgroups but not others.

A final qualification comes from the finding that one of the program impacts for the observational measures of literacy interactions was in an unexpected direction. Mothers in the experimental group had significantly *lower* scores on the Ease of Ideas measure during the wheels task, which involved the mother eliciting from the child the names of objects with wheels. Upon closer examination, this group difference was found to be attributable entirely to a higher proportion of mothers in the experimental group who failed to understand that they should provide clues to the child that would help him or her name wheeled objects.

As noted in Chapter 5, interpretation of this negative finding must remain open at this point. It is possible, for example, that interviewer instructions were less complete or clear for those mothers who did not grasp the task and that by chance such unclear instructions occurred more often for mothers in the experimental group.[1] It is also possible that mothers in New Chance, because of their participation in educational programs, have been sensitized to testing situations. Perhaps the wheels task, which involved no physical props and was left entirely to the mother to structure, felt like a test and caused anxiety in some mothers, particularly those in the experimental group. Such an interpretation would not suggest that New Chance had important negative implications for mother-child literacy interactions under everyday circumstances, but rather implications for mothers' test anxiety. A final interpretation would suggest that participation in New Chance does indeed have negative implications for certain kinds of mother-child literacy interactions. This interpretation appears unwarranted, however, in light of the fact that no group difference in Ease of Ideas was found on the wheels task once the analysis was limited to those mothers who grasped the purpose of the task. It will be important in future work to attempt to distinguish among the possibilities that the wheels task itself is problematic for a population like that participating in New Chance, perhaps eliciting test anxiety; that instructions for the task (or their presentation by the interviewers) are problematic and need clarification; or that a program such as New Chance may have unintended negative effects on mothers' interactions with their children on tasks involving guessing and categorization. Because the finding for Ease of Ideas does not unambiguously reflect a program impact on mother-child literacy interactions and appears open to several interpretations at this point, we focus in the remainder of this chapter on the other documented program impacts.

---

[1] Part II of this monograph notes that interviewer effects are a possible, but rarely studied, source of measurement error. Interviewer effects could have contributed to the pattern noted here if (1) certain interviewers were assigned unevenly across the experimental and control groups; and (2) some interviewers assigned more often to carry out the observational study with families from the experimental group were less skilled in introducing this task or confirming that the mother understood it.

B. Magnitude of Program Impacts on Parenting

We conclude, to this point, that findings of positive program impacts on parenting appear across parenting measures with very different characteristics. Yet we note that within types of parenting measures, significant impacts are not pervasive and, further, that group differences in most instances do not occur for all population subgroups. Looking across the differing measures of parenting, we next consider the magnitude of positive program effects, asking whether stronger effects were documented for particular measures than for others.

For each of the measures of parenting on which positive program impacts have been found, we can examine the magnitude of the impact by computing effect sizes. Effect sizes provide a standardized measure (in standard deviation units) of the strength of the impact. According to Cohen (1988) an effect size of .2 (or one-fifth of a standard deviation) can be interpreted as a small effect; .5 (or one-half of a standard deviation) as a medium effect; and .8 (or four-fifths of a standard deviation) as a large effect.

We have examined effect sizes for five of the seven parenting variables for which significant positive program impacts were documented. In those instances where there were significant program impacts for both a total score and for a subscale that was one component of a total score, we restricted our focus to the more specific measure (that is, we examined effect sizes for the program impact on the HOME Emotional Support subscale, but not the HOME total; and for Parenting Chore Time, but not Overall Parenting Time). Thus, we examined effect sizes for the following parenting measures: Book Reading Quality, Harsh Treatment, Maternal Warmth, Emotional Support, and Parenting Chore Time.

The effect size for the HOME Emotional Support subscale fell in the medium range, d = .50. For the remaining four variables with significant impacts, effect sizes fell in the small-to-medium range. Specifically, effect sizes were .31 for Parenting Chore Time, .27 for Harsh Treatment, .26 for Maternal Warmth, and .23 for Book Reading Quality. While these effect sizes fall in the small-to-medium range, we note that some program evaluations have viewed effect sizes within this range as having policy-relevance (for example, the JOBS Child Outcomes Study; Moore et al., 1995).

C. Identifying Program Components Associated with the Parenting Impacts

Chase-Lansdale, Brooks-Gunn, and Paikoff (1991) have urged that evaluations of programs for adolescent mothers seek to document *which* components of comprehensive programs contribute to positive impacts on parenting behavior. We turn now to analyses exploring this issue for the New Chance Observational Study.

In Chapter 3 we articulated four *a priori* hypotheses for different ways in which the New Chance Program could affect parenting: through (1) participation specifically in the parenting education component of the program; (2) participation in the wider set of program components that addressed parenting behavior in some way; (3) participation in the human capital development components of the program, aimed at equipping the mother with better educational and employment skills; and (4) overall contact with the program. We have examined each of these aspects of program participation in relation to the parenting measures, first using a "between-groups" approach, and then looking within the experimental group alone.

The between-groups approach involved asking whether the experimental-control group difference on any of the parenting variables was *significantly diminished* by controlling for program participation variables individually and in the noted combinations (in addition to the standard set of control variables included in all impact analyses). These analyses relied on the maternal report measures of program participation, as the more precise Management Information System (MIS) data are available only within the experimental group. In these analyses, none of the measured aspects of program participation, alone or in combination, accounted for the group difference on any of the measures of parenting behavior. The possibility remains that aspects of program participation not captured by these participation variables may account for the group differences in parenting behavior.

Our next step was to use an approach focusing on the experimental group only. In the report of the 18-month follow-up survey in the larger New Chance Evaluation, Quint and colleagues (1994) followed a "within-group" approach to examining the program components underlying impacts on the parenting measures included in that survey. Their strategy asked whether, within the group of mothers assigned to the experimental group, extent of participation was associated with parenting outcomes. Level of participation in parenting education classes, and in the New Chance Program overall, was divided into no participation, lower, middle, and upper third of participation (hereinafter referred to as none, low, moderate and high participation categories), using the MIS data. The program impact on the HOME-SF, documented for the full evaluation sample, was found to be related to total program participation rather than participation in parenting education when defined in this way.

The authors of the 18-month report note the need for caution in interpreting these results: those experimental group mothers who choose to participate *more* in the New Chance Program, as opposed to *less*, may be a select group of mothers. This group may already be showing higher quality parenting behavior. That is, some underlying characteristic or set of characteristics (for example, related to the social support available to the young mother at home) may serve as the basis of both program participation and positive parenting behavior. Thus, self-selection *within* the experimental group may explain the association between program participation and the parenting impact.

We sought to extend this set of analyses from the 18-month report in several ways (although, of course, focusing here only on the New Chance Observational Study sample): (1) by examining all of the variables for the observational study sample on which we have documented parenting effects (a larger set of variables than for the full evaluation sample); (2) by including consideration not only of participation in parenting education and total program participation as aspects of participation that may help explain impacts on parenting variables, but also the intermediary points of participation in the set of key program components that had some content relevant to parenting, and participation in the human capital development aspects of the program; and (3) by attempting to take into account in our analyses those background characteristics related to program participation.

In Chapter 3 we noted that MIS data could be used to create not only a summary score of hours of participation in parenting education classes and in the New Chance Program overall, but also a proxy summary score that represents participation in a set of "key" program components with content related in some way to parenting behavior, and a summary score of participation in human capital development program components. We noted that the key components summary score was an approximation of the desired summary score in that data regarding two components hypothesized to be related to parenting, use of child care and personal counseling, could not be obtained from the MIS data. Chapter 3 also reported on the set of background characteristics significantly related to, or selecting for, extent of participation defined in each of the four ways we will examine here.

Table 7.1 shows the findings of our within-group analyses. These analyses categorize level of participation as none, low, moderate, or high, in terms of MIS data for hours of participation. We focus on the same five parenting measures as in the examination of effect sizes (again choosing to focus on specific subscales where impacts had occurred for both a subscale and a total or summary measure).

A first set of analyses of variance asks whether means on these five parenting variables differed significantly according to level of participation (defined according to each of the four *a priori* hypotheses). Then analyses of covariance repeat these analyses, controlling for those baseline characteristics significantly associated with program participation. The covariates included differed according to the hypothesis being examined (that is, different background characteristics were significantly associated with participation in parenting education alone, participation in the key components, participation in the human capital development components, and total program participation, as discussed in Chapter 3).

The p-values in the top row associated with each hypothesis in Table 7.1 show that only three of the five parenting variables — Harsh Treatment, Book Reading Quality, and Emotional Support — differed significantly in light of level of participation as defined in any of the four ways. For each of these three, significant differences in mean levels of parenting were observed according to participation in the human capital development program components and in the program as a whole. In addition, the Emotional Support subscale also differed significantly according to both level of participation in the key components involving some content relevant to parenting, and according to participation specifically in parenting education classes.

In the "Means Adjusted" part of Table 7.1, we controlled for those baseline variables that were significantly related to participation as defined in each of the four ways (see Chapter 3 for a list of these baseline variables). We note that there may have been further background characteristics, not documented in the baseline instrument, ("nonobservables"), which could not be controlled.

The "adjusted" analyses reveal that means for the Emotional Support subscale continued to differ significantly according to participation as defined in each of the four ways. Table 7.1 shows that within the experimental group, scores on the Emotional Support subscale were highest for mothers who had participated the most in parenting education, in the key program components with some content relevant to parenting, in the human capital development components of

Table 7.1

Within-Group Analyses
Means on Parenting Outcomes, by Program Participation

	Parenting Outcomes: Nonadjusted Analyses				
Alternative Ways for New Chance to Affect Parenting	Harsh Treatment	Book Reading Quality	HOME Emotional Support	Maternal Warmth	Parenting Chore Time
Hypothesis I: Via Participation in Parenting Classes	p = .507	p = .546	p = .059 *	p = .383	p = .995
None	.25	6.59	95.47	25.49	1.33
Low	.44	6.58	97.20	23.68	1.28
Moderate	.23	6.56	98.91	24.33	1.26
High	.28	6.96	103.88	23.22	1.27
Hypothesis II: Via Participation in Key Components	p = .773	p = .632	p = .003 ***	p = .695	p = .893
None	.32	6.57	99.21	24.84	1.39
Low	.34	6.48	93.08	24.34	1.31
Moderate	.19	6.74	100.21	23.26	1.33
High	.29	6.88	104.26	24.30	1.17
Hypothesis III: Via Participation in Human Capital Development	p = .068 *	p = .002 ***	p = .001 ***	p = .739	p = .384
None	.31	6.57	98.21	25.04	1.48
Low	.28	6.29	93.07	24.33	1.30
Moderate	.48	6.40	99.80	24.00	1.42
High	.08	7.35	105.38	23.44	1.03
Hypothesis IV: Via Total Participation	p = .093 *	p = .004 ***	p = .000 ***	p = .793	p = .809
None	.36	6.80	97.84	25.04	1.27
Low	.28	6.20	93.61	24.11	1.38
Moderate	.46	6.45	98.06	24.25	1.34
High	.09	7.25	106.38	23.51	1.14

(continued)

Table 7.1 (continued)

Alternative Ways for New Chance to Affect Parenting	Parenting Outcomes: Means Adjusted for Selection Factors[a] Harsh Treatment	Book Reading Quality	HOME Emotional Support	Maternal Warmth	Parenting Chore Time
Hypothesis I: Via Participation in Parenting Classes	p = .548	p = .898	p = .089 *	p = .502	p = .971
None	.20	6.58	95.70	25.44	1.37
Low	.43	6.68	98.34	24.08	1.27
Moderate	.24	6.58	98.81	24.32	1.27
High	.25	6.80	104.01	23.34	1.23
Hypothesis II: Via Participation in Key Components	p = .903	p = .869	p = .018 **	p = .853	p = .772
None	.29	6.60	100.53	24.71	1.44
Low	.31	6.55	94.52	24.49	1.28
Moderate	.21	6.66	99.39	23.59	1.35
High	.31	6.81	103.91	24.30	1.13
Hypothesis III: Via Participation in Human Capital Development	p = .123	p = .005 ***	p = .001 ***	p = .882	p = .265
None	.28	6.59	98.56	24.85	1.52
Low	.28	6.30	93.12	24.23	1.32
Moderate	.47	6.41	99.90	23.97	1.42
High	.11	7.32	105.06	23.67	.98
Hypothesis IV: Via Total Participation	p = .213	p = .010 **	p = .001 ***	p = .906	p = .564
None	.33	6.86	99.01	24.97	1.33
Low	.27	6.22	93.98	24.01	1.42
Moderate	.45	6.45	97.88	24.10	1.35
High	.13	7.21	105.73	23.78	1.07

Sources: New Chance 18-month follow-up survey, brief interview accompanying the observational session, coded observational study variables, and New Chance Management Information System (MIS) data.

Notes: Sample was restricted to experimental group only, n = 184. The sample size may fall short of this number because of missing or unusable items from some respondents' questionnaires. Statistical significance levels are indicated as *** <= 1 percent, ** <= 5 percent, and * < =10 percent, which refers to the probability that means are different from each other only because of random error.

[a] Means were statistically adjusted for effects of baseline characteristics that were found to be associated with extent of participation (see Chapter 3).

the program, and in the program overall. The adjusted analyses also reveal that means for the observational Book Reading Quality measure continued to differ significantly according to participation in the human capital development components of the program and according to overall program participation. The highest mean scores on Book Reading Quality were obtained by those who had high levels of participation in the human capital development program components and the program overall.[2]

Where significant differences across groups were identified in both the adjusted and nonadjusted analyses, post hoc mean comparisons[3] among nonadjusted means were carried out to locate the pairs of means (no participation, low, moderate, and high participation) that differed significantly. In all instances except one in Table 7.1, where an overall group difference was documented for adjusted means, significant differences occurred between the *low* and *high* participation groups. For the "parenting classes" hypothesis, no significant differences emerged among the nonadjusted means for Emotional Support. In addition, for the "total participation" hypothesis the *moderate* and *high* groups differed for the Emotional Support variables. For the "human capital development" hypothesis, a difference occurred between the *moderate* and *high* groups for Book Reading Quality. It is interesting to note that no differences were found between those who did not participate at all and other groups. Examination of the means reveals that across variables it was most often those with *low* participation, rather than those with *no* participation, who had the lowest means on parenting variables. One might speculate that some of the mothers opting not to participate in various program components accurately perceived themselves as less in need of personal development.

These analyses provide some modest support for the hypothesis that participation in the parenting education component of the New Chance Program helped to improve parenting behavior. But the findings also suggest that other aspects of the program, even aspects with no content directly related to parenting, had the potential to affect parenting behavior. It may be important for those working with adolescent mothers to consider the possibility that enhancing mothers' education and employment skills may simultaneously enhance their parenting behaviors. Those formulating programs for adolescent mothers should also consider the possibility that comprehensive programs, perhaps through the overall support and encouragement they provide, can have positive implications for parenting behavior.

We must, however, point to caveats in the interpretation of these findings. The between-groups analyses that were carried out shed no light on the issue of program components related to parenting impacts. Further, the within-group analyses point to no linkages between program participation and two of the parenting measures for which there were impacts: Maternal Warmth and Parenting Chore Time. Our measure of participation in key program components could not reflect participation in child care and personal counseling, two program components that we hypothesized to have the potential to affect parenting. Finally, while we have attempted in our within-group analyses to control for self-selection, the possibility remains that further undocumented background characteristics encouraged both program participation and more supportive and stimulating parenting.

---

[2] The possibility that these findings could have arisen due to outlier cases on participation was reviewed. No outliers were found on participation as defined by Hypotheses II, III, and IV. One outlier case was found for Hypothesis I. However, because with selection factors controlled, differences for Hypothesis I are no longer significant, we did not repeat this analysis deleting the outlier case.

[3] The procedure used for post hoc mean comparisons was the Scheffe test, widely considered to be a conservative estimate of differences among means.

Keeping these caveats in mind, future work should explore in greater detail the possibility that in a comprehensive program such as New Chance, parenting impacts can grow out of program components aimed at mothers' human capital development, and from a synergy across program components, and not only out of components aimed specifically at parenting behavior. That participation in the parenting education component of the New Chance Program was not more widely associated with parenting behavior may reflect the need for a greater dosage of parenting education in a program for adolescent mothers, and the need for such parenting education programs to address explicitly the obstacles that young mothers face to changing their parenting behavior.

II. The Contribution of Observational Measures of Parenting to the New Chance Evaluation

What have we added to the evaluation of New Chance by adding the observational measures to the approaches already available within the larger evaluation for measuring parenting behavior? How can we look across the differing types of parenting measures, examining in particular the unique contribution made by the observational measures?

In this section we take three different approaches to addressing these questions. First, we look at the interview-based and observational measures of parenting conceptually, examining them in terms of the intent and content of each. In the monograph introduction we noted a starting assumption that by their nature, the observational measures differed from the survey measures of parenting, and could be viewed as providing information to complement that derived from the interview-based measures. Is this assumption borne out by a careful contrast of the nature of each measurement approach?

Second, we examine the correlations among the different parenting measures. A key issue is that of the extent to which the observational measures are correlated with the other kinds of parenting measures. Consistently high correlations of the observational measure of parenting and the interview-based measures (that is, the HOME-SF subscales, which rely on maternal report and interviewer ratings of what is observed during the interview; and the Maternal Report Parenting scales and Time Use measures, which rely entirely on maternal report) would support the conclusion that the observational measures overlap to a great extent with the other types of parenting measures and contribute little that is unique. A lack of significant correlations, or a pattern of low to moderate correlations, would indicate that the measures are substantively different.

Third, we ask how the different kinds of parenting measures function as predictors of child outcomes at the 42-month follow-up within the observational study sample. Linking parenting measures to child outcomes is a critical marker of the utility of differing parenting measures that has been used in previous research (Berlin et al., 1995; Berlin and Zaslow, 1996). Here we ask whether the interview-based and observational measures of parenting *each* serve as significant predictors of selected child outcomes once family background characteristics are taken into account. As a stringent examination of the role of observational measures in particular, we then ask whether we improve our ability to predict to child outcomes when the observational measures are added to the other parenting measures in predicting the child outcomes.

A. Conceptual Evaluation: The Intent and Content of the Differing Parenting Measures

The parenting measures included in the New Chance Observational Study differ in three respects: in terms of *informant*, or who provided the data about parenting behavior; *scope* of information collected by each type of measure; and *content*, or aspects of parenting considered. Table 7.2 summarizes key differences across the measures in these respects. The observational measures differ from the other types of parenting measures particularly with regard to informant and scope, although there are some notable differences with regard to content as well. In the section that follows we highlight key differences across these measures.

1. **Informant.** As can be seen in Table 7.2, mothers themselves serve as informants to at least some degree for each type of parenting measure except for the observational measures. Mothers are the only appropriate informant for some of the measures noted in the table. For example, the measures involving maternal attitudes or beliefs (such as the measure of Parenting Stress) and the summary of time allocated to parenting activities during the previous day must rely on maternal report.

However, researchers have repeatedly raised concerns about maternal report as a source of information on the occurrence or quality of particular parenting *behaviors*. For example, in reviewing the history of the HOME Inventory, Elardo and Bradley (1981, p. 117) note:

> It was Caldwell's opinion that interviewing techniques were generally not adequate to assess such critical parent behaviors as responsivity and warmth. . . . In addition, interview techniques are limited due to their exclusive reliance on parental reports — reports which may suffer from both inaccuracy of recall and dishonesty of response. . . . Therefore, a combination of observation and interview techniques was chosen for the HOME scale.

While Elardo and Bradley call attention to problems of bias and inaccuracy of recall in maternal report, it is also possible that for certain aspects of parenting behavior, mothers themselves are simply limited as informants. Parenting constructs that are important to document (for example the ratings in the observational study of Supportive Presence or Number of Nonimmediate Utterances) may not be ones that mothers think in terms of at all. Mothers may lack an awareness of how their own behavior compares with that of other mothers, though such a comparative framework may be crucial to providing a rating. Further, mothers may lack the self-reflection or insight that would be necessary to report accurately on their own behavior in terms of these constructs.

In reporting on the findings of the New Chance 18-month survey, Quint and colleagues (1994) make special note that the Emotional Support subscale of the HOME-SF for which a (small but significant) program impact was identified in the full evaluation sample relies heavily on interviewer ratings. They consider this finding therefore more likely to be free of response

Table 7.2

Comparing the Four Types of Parenting Measures Included in the New Chance Observational Study

	Parenting Measures			
	Collected at 18-month follow-up		***Collected around 21 months***	
	HOME - SF	**Mother Report**	**Time Use**	**Observational**
	Emotional Support	**Stress**	**Overall Parenting Time**	**Literacy-Related Behaviors**
	Cognitive Stimulation	**Control**	**Parenting Chores**	**Affective Quality**
	Physical Environment	**Warmth**		
	Harsh Discipline			
Informant				
Identity	Mother Interviewer	Mother	Mother	Coder
Extent of Training for Informant	Some training of interviewer required for interviewer ratings	None	None	Extensive and ongoing
Interrater reliability obtained where informant is not mother?	No	—	—	Yes
Informant blind as to research group?	No	No	No	Yes

(continued)

Table 7.2 (continued)

	Parenting Measures			
	Collected at 18-month follow-up		***Collected around 21 months***	
	HOME - SF	**Mother Report**	**Time Use**	**Observational**
Scope				
Period of time to which ratings or questions pertain?	Period of interviewer's presence in the home, a week, month, year, or open, depending on item	Open	Single specific weekday	Duration of mother-child task or tasks
Focus exclusively on mother-child dyad?	No Father, other family members also included	Mostly Some items concern mothers and children in general	No Time spent on parenting includes other children	Yes
Broader physical environment considered?	Yes; physical features of home, materials for cognitive stimulation available to child, and to some extent area immediately outside of home	Items describe interactions without reference to where they occur	Documents activities over course of full day without reference to where they occur	No; provision of standard set of task materials Features of larger environment rated only from point of view of interruptions to tasks

(continued)

Table 7.2 (continued)

	Parenting Measures			
	Collected at 18-month follow-up		*Collected around 21 months*	
	HOME - SF	**Mother Report**	**Time Use**	**Observational**
Nature of Behavior Considered				
Cognitive and/or socioemotional aspects of interaction considered?	Cognitive and socioemotional	Socioemotional		Cognitive and socio-emotional
Ongoing behavior or behavior during challenge?	Ongoing	Ongoing and when challenged	Ongoing	Challenge
Focus on part of behavioral continuum or full range from low to high?	Ratings by mother and interviewer summarized in terms of whether a feature of the home or of interactions is or is not in a negative range	Full range	Full range	Full range
Focus is primarily on qualitative aspects of behavior, or on occurrence/non-occurrence/frequency	Occurrence/ frequency	Quality of interactions	Occurrence/ frequency	Quality of interactions
Ratings involve behavioral distinctions that are subtle, fine-grained and require training to make, or rely on, commonly used categories of behavior	Commonly used categories of behavior	Commonly used categories of behavior	Commonly used categories of behavior	Most ratings involve subtle and fine-grained distinctions

bias than one based on a measure relying heavily on maternal report. However, in the preceding chapter we noted the possible sources of bias in interviewer ratings as well (for example, awareness of a great deal of further information through the interview itself). Thus, though interviewer ratings may be less subject to response biases than maternal report measures, they are not free of them.

The fact that the observational measures do not rely at all on maternal report or interviewer ratings strengthens the overall evaluation by providing data free of biases involved in collecting information in an interview context, and by permitting the collection of data on aspects of behavior about which mothers may not be aware. The observational measures are also the only set of parenting measures for which the informant was "blind" (completely unaware) as to the group that the family had been randomly assigned to within the evaluation. Interviewers for the full evaluation not only were aware of families' assigned research groups, but also were required to ask for details of educational activities, employment, and child care. We have very little information about whether interviewers had any hypotheses about how the New Chance Intervention or any of its components affected parenting behavior.[4] Any such assumptions about the intervention could have been a further source of bias in the HOME-SF as well. Further, mothers' sense of how program participation was intended to affect parenting could have affected the Maternal Report Parenting scales and Time Use measures. For all of these reasons the observational measures provide an important resource for the larger evaluation, in helping to confirm that differences in parenting behavior occur.

Table 7.2 notes further that for the three sets of measures involving an outside observer (the interviewer for the HOME-SF and the coder for the two sets of observational measures — literacy-related and concerning affective quality), a clear difference exists with regard to extensiveness of training required in order to complete behavioral ratings. In order to administer the full HOME Inventory, a home visitor must first demonstrate that her ratings reach a high level of agreement with an experienced coder or with the ratings for a videotaped interview that serves as a standard. Interrater agreement is ascertained not only before an interviewer is considered qualified, but also as the study proceeds (see, for example, Bradley et al., 1989, 1994). The HOME-SF, which was employed in the New Chance Evaluation, is generally administered by survey interviewers who receive item-by-item training during a training conference lasting two or three days, during which many topics are covered. Survey interviewers administering the HOME-SF typically are not required to demonstrate agreement with an experienced coder or training videotape (Baker and Mott, 1989). By contrast, the observational measures require extensive initial training and repeated reassessment over time of interrater reliability.

In sum, the fact that the observational measures are free of biases associated with maternal report, that coders for the observational study are unaware of research group, and that training of coders for the observational measures is extensive, rigorous, and ongoing permits us to view the observational measures as a valuable resource to the overall evaluation and strengthens our confidence in the differences found using measures without these safeguards. At the same time, it is not possible to argue that the observational measures are entirely free of the possibility of bias.

---

[4] The distance between survey interviewers and researchers (discussed in Chapter 10) would work against interviewers being aware of hypotheses already articulated by the researchers.

In particular, participation in New Chance may have alerted mothers to which interactive behaviors were viewed as positive and which as negative to show during the mother-child tasks. Group differences documented from the videotaped interactions may reflect not so much the nature of *ongoing* mother-child interaction as interaction in which mothers are seeking to demonstrate, with "cameras rolling," the behaviors that they have learned are valued. Further, the lack of match in terms of racial/ethnic identity or social class between those who coded the videotapes and transcripts and the families in the New Chance Observational Study may have introduced bias, albeit one that applied equally to families in the experimental and control groups.

2. Scope. Table 7.2 suggests that by contrast with the other parenting measures, and particularly the HOME-SF subscales, the observational measures are focused rather than broad in scope. Using the metaphor of windows into a house (ways of looking into the terrain of parent-child relations), the observational measures appear to shed an intense light through a fairly small and defined window. By contrast, other measures provide a larger window, but with perhaps a less intense light. From this perspective, the observational study measures appear to complement the approach taken by the other parenting measures.

The HOME-SF subscales emerge as particularly broad in scope with regard to time frame considered, individuals whose input into the child's development is considered, and extent to which the larger physical environment is taken into account. While some of the HOME-SF questions and ratings pertain only to the period that the interviewer was in the home, other items consider periods of weeks (for example, asking whether a family member has taken the child on an outing at least every other week) or the mother's parenting practices without regard to time period (for example, asking whether the child is permitted to hit the parent without reprisal). In addition to considering the mother's input into the child's development, the HOME-SF also asks about contact and activities with the father (for example, a question regarding whether the child eats at least one meal per day, on most days, with mother and father), and about outings with other family members. Together the mother and interviewer provide a great deal of information about the provision of cognitive stimulation through the physical environment of the home and the family as a whole.

By contrast, the particular observational procedures used in the New Chance Observational Study focus *only* on the time period during which the mother and child are carrying out the tasks and *only* on the mother-child dyad. Further, mothers and children are asked to attend *only* to the particular set of materials provided by the researchers. While this holds constant the social context and the physical stimuli that provide the basis of ratings, it also precludes consideration of the contribution of other material and social sources of stimulation and support available to the child in the household.

The Maternal Report Parenting scales and Time Use measures appear to take intermediate positions with regard to scope. For example, looking at time frame, a progression can be seen from the immediate focus of the observational measures, the focus on a single day of the Time Use data, and the less time-bound focus of the Maternal Report Parenting scales and the HOME-SF.

These differences suggest a contrast in the intent of the measures. At one extreme, the

HOME-SF appears to seek the global input of significant adults and the physical environment of the home to the child's stimulation and support in everyday life. The observational measures, by contrast, seek to capture, much more specifically, the quality of the mother's interactions with the child in a specific and delimited situation: when the dyad is taxed with challenging tasks, the assumption being that behavior in this context is reflective of critical aspects of the relationship.

3. Content. While the observational measures and the HOME-SF subscales appear to be different with regard to informant and scope, they are similar with regard to the domains of parenting behavior each seeks to address. Both make a basic distinction between the socioemotional aspects of parenting behavior and those involved in the provision of cognitive stimulation, and both provide measures regarding each aspect. The particular content of the socioemotional components of the HOME-SF and the observational measures agree quite closely, each focusing on warmth and support in the relationship as well as on issues of discipline. The Maternal Report Parenting scales, too, address these socioemotional aspects of parenting. The HOME-SF and observational measures of cognitive stimulation in this study correspond less closely, the latter focusing specifically on literacy-related aspects of mother-child interaction and the former considering multiple types of cognitive stimulation (for example, from toys in the home, the organization and perceptual characteristics of the household, outings, and direct mother-child interactions).

From the point of view of particular parenting behaviors considered, the Time Use measures stand apart. These measures do not seek to document the quality of interactions, but rather the sheer allocation of the mother's time spent with the child or in caregiving activities with the child. Time use studies have proven important in understanding other issues in children's development. For example, studies of children's time use document that during adolescence time spent with parents decreases while time spent in interactions with friends increases (Savin-Williams and Berndt, 1990). Time use data have similarly proven useful in documenting differences between employed and nonemployed mothers in both the quantity of time spent with children and the nature of activities engaged in during time together (Hill and Stafford, 1979).

While the HOME-SF, Maternal Report Parenting scales, and observational measures show similarity in terms of the aspects of parenting they focus on, they differ in terms of the circumstances they sample to obtain a description of these behaviors. Berlin and colleagues (1995), in contrasting observational measures and the HOME Inventory, liken the difference to that between measuring heart functioning during normal activity and measuring it in the context of a stress test. The HOME Inventory seeks to describe the cognitive stimulation and emotional support available to the child during normal home-based activities. The observational measures capture maternal and child behavior during the mild stress of challenging tasks.

The challenge context assumes that certain maternal and child behaviors are particularly important to capture in order to characterize the relationship, and these behaviors are best captured when the child must turn to the mother for help and support (Egeland et al., 1995). A session sampling ongoing activities might not provide an opportunity to view the mother's sensitivity and effectiveness in guiding the child through a difficult task or her tendency to encourage or rebuke the child's efforts where success is not readily achieved. Her effectiveness in such a circumstance and the child's openness to and reliance on her input may be key descriptors of the relationship. The challenging task context of the observational study in this sense resembles the mild stress carefully orchestrated for assessments of child attachment behavior (Sroufe, 1979). As shown in Table 7.2, the Time Use measures reflect on maternal behavior during normal activity (which we assume to encompass both stressful and nonstressful occurrences, taking place spontaneously rather than in an orchestrated manner). The Maternal Report Parenting scales encompass both the perspectives of normal activity and items pertaining specifically to stressful circumstances (for example, "Even when I'm in a bad mood, I show my child a lot of love").

Just as a thorough examination of heart functioning requires testing during normal activity and during stress, we view both the measures of parenting that tap interactions during ongoing activity and those that record behavior during challenging circumstances as contributing in important ways to the overall portrayal. The fact that numerous studies of parenting, including several focusing on adolescent mothers, have included both of these types of measures side by side (for example, Crockenberg, 1979; Garcia-Coll, Hoffman, and Oh, 1987; Schilmoeller and Baranowski, 1985; Unger and Wandersman, 1988; Walker, et al., 1995) implicitly supports the assumption that these are complementary approaches.

Several further distinctions can be made in terms of how cognitive stimulation and emotional support are recorded in the differing parenting measures. We have already noted that the HOME-SF tends to truncate the full behavioral continuum because it requires a decision as to whether a particular behavior or environmental condition involves risk to the child. For example, rather than focusing on the full range of possible disciplinary techniques, the HOME-SF summary rating deals primarily with whether or not physical punishment has been used. Each of the remaining parenting measures can be characterized as "full continuum" measures, providing ratings from the low end to the high end of each parenting behavior considered.

A further distinction regarding the content of ratings is that of whether each type of measure focuses primarily on the *occurrence or frequency* of predefined behaviors or on the *quality* of a behavior when it is occurring. This distinction may best be illustrated by considering the observational measures of literacy. The measures rated by Snow and colleagues from the videotapes, while encompassing several measures of frequency (for example, Number of Nonimmediate Utterances), also focus extensively on the quality of mother-child literacy interactions (for example, Book Reading Quality, Ease of Ideas during the wheels task). By contrast, the cognitive stimulation items in the HOME-SF focus exclusively on quantities (for example, the number of toys and books in the home; the frequency of book reading or outings).

4. Summary. The different types of parenting measures have different goals and strengths. The observational measures shed an intense light on the quality of interactions within the mother-child dyad under conditions of challenge. They rely upon coders who are unaware of research group or further information about the family and who are extensively and rigorously trained. The categories of behavior recorded through the observational research are critical to understanding the experiences of the child, and yet are not aspects of the relationship on which mothers themselves or even trained interviewers could be expected to reflect.

A strength of the HOME-SF is its breadth of scope, capturing not only aspects of the mother-child relationship, but also input from the physical environment of the home and from other social partners. Unlike the observational measures, which seek to record characteristics of interaction ranging from extremely negative to extremely positive, the HOME Inventory is a measure of environmental risk and provides a summary of the number of circumstances of daily life that can be seen as posing barriers to children's development. This measure considers everyday life rather than seeking to distill out how mothers and children turn to each other when challenged.

The Maternal Report Parenting scales further complete the picture of parenting behavior by tapping mothers' subjective reactions and attitudes. The Parenting Stress measure, for example, helps us to understand whether participation in a program like New Chance alters mothers' subjective experience of daily aggravation in the parenting role. It is important to be able to reach conclusions not only about externally apparent aspects of parenting, but also about mothers' internal reactions.

Finally, the Time Use variables do not probe the quality of the child's experiences either with the mother or in the home in general, or the mother's reactions to the parenting role, but rather her allocation of time to this role. It is possible that a program like New Chance may change the mother's commitment to parenting in the sense of her allocation of time to this role even in the absence of changes in the quality of her interactions or her subjective sense of parenting. Alternately, a program like New Chance has the potential of diminishing the time a mother has available to the parenting role because it increases her involvement in activities outside the home. The inclusion of Time Use variables helps separate the quality and quantity of time spent in parenting activities.

We conclude that the different types of parenting measures provide differing and complementary perspectives on the parent-child relationship, all of them useful in an evaluation such as New Chance.

B. Correlations Among the Different Types of Parenting Measures

How do the correlations across the different types of parenting measures correspond to our portrayal? Is it the case that the similarities in content covered, particularly between the observational measures and the HOME-SF subscales, result in such high correlations that we cannot consider these distinct measures? Further, are there significant correlations where one would seek them — for example, between the observational measure of Harsh Treatment and the HOME-SF and Maternal Report measures of maternal discipline?

Table 7.3 presents the correlations of the observational measures of parenting with the HOME-SF subscales, the Maternal Report Parenting scales, and the Time Use measures. We have followed the precedent of Chapters 4 and 5 of selecting specific observational measures to focus upon rather than considering the full set of observational measures coded by the University of Minnesota and Harvard University teams.

In the sections that follow we ask a series of specific questions to facilitate the task of navigating through the large number of data points in a table of correlations. We focus on each block of correlations in the table separately. A "block" is defined as the set of correlations between *one* group of observational measures (the five measures of affective quality, the five

Table 7.3

Correlations Among Various Parenting Measures

	Interview-Based Measures									
	HOME-SF					Maternal Report			Time Use	
Observational Measure	Emotional Support	Cognitive Stimulation	Physical Environment	Harsh Discipline[b]	Total	Stress	Control	Warmth	Overall Parenting Time	Parenting Chores Time
Affective Quality: All Tasks										
Mother's Supportive Presence	.16 *	.24 **	.11	.14 *	.27 **	-.06	-.25 **	.02	.04	.07
Mother's Harsh Treatment[a]	-.13 *	-.20 **	-.09	-.13 *	-.23 **	.10	.14 *	.01	-.13 *	-.13 *
Child's Compliance	.05	.06	.03	.02	.08	.04	-.01	.01	.03	-.02
Child's Affection to Mother	.15 *	.11	.19 **	.05	.23 **	-.03	-.09	.01	.13 *	.12
Quality of Relationship	.17 **	.19 **	.17 *	.08	.27 **	-.01	-.12	-.01	.08	.11
Literacy: Book Reading Task										
Total Number of Utterances	-.05	.04	.06	-.06	-.01	.04	-.03	.08	.05	-.01
Number of Nonimmediate Utterances	.09	.12	.09	-.04	.13	.04	.02	-.01	.01	.05
Percentage of Immediate Utterances	-.01	-.07	-.04	.03	-.06	-.10	.04	.06	.03	-.03
Book Reading Quality	.22 **	.27 **	.16 *	.05	.29 **	-.01	-.27 **	.05	.05	-.02
Number of Discussion Topics	.02	.01	.06	-.07	.03	-.07	-.17 **	-.07	.06	-.02
Literacy: Wheels Task										
Objects Named	.05	.15 *	.11	-.07	.15 *	-.05	-.05	-.05	.07	.01
Objects/Elicitations	.01	.14 *	.04	-.04	.09	-.10	-.04	-.07	.09	.10
Mother's Ease of Ideas	-.01	.18 **	.03	-.02	.11	-.07	-.10	-.03	.03	-.01

Sources: New Chance 18-month follow-up survey, brief interview accompanying the observational session, and coded observational study variables.

Notes: Sample size was 229 owing to listwise deletion of cases. Statistical significance levels are indicated as *** <= 1 percent, ** <=5 percent, and * <= 10 percent.

[a] Higher scores indicate more harsh treatment.

[b] Higher scores indicated less harsh discipline.

measures of the book reading task, *or* the three measures of the wheels task); and *one* group of interview-based measures (the five HOME-SF scales, the three Maternal Report Parenting scales, *or* the two Time Use measures). The block appearing in the upper left-hand corner of Table 7.3, for example, consists of the correlations between the observational measures of affective quality and the HOME-SF subscales. Table 7.4 summarizes the correlations within each block. This table notes the range of correlations within a block, the proportion of correlations within the block that was statistically significant, and the variables that were most highly correlated within the block.

1. **Correlations Between Observational and Interview-Based Measures.** Tables 7.3 and 7.4 quickly yield the conclusion that correlations between the observational and interview-based parenting measures are not high. Although a fair number of correlations are significant, the correlations never exceed .29. It is not the case, then, that the observational measures of parenting are so highly correlated with the other parenting measures as to be considered largely overlapping.

2. **Variables Showing Highest Correlations with the Observational Measures.** The summary measures shown in Table 7.4 indicate that the highest correlations and the highest proportions of significant correlations occur within the blocks relating the observational measures to the HOME-SF subscales. By contrast, correlations of the observational measures with the three Maternal Report Parenting scales (Warmth, Control, and Stress) or the two Time Use variables (Overall Parenting Time and Parenting Chore Time) are low and rarely significant. There are no significant correlations at all between the observational measures of literacy and the Time Use measures.

This summary of the pattern of correlations corresponds to our conceptual evaluation of the content of parenting measures, in which we noted that the observational measures corresponded most closely in content to the HOME-SF subscales, and least closely to the Time Use measures. However, the Maternal Parenting scales are also related in content to the observational measures, particularly those focusing on the affective quality of mother-child interaction. The general lack of significant correlations here is somewhat surprising.

While in general the observational measures and the HOME-SF subscales are most closely related, there are some noteworthy differences when the observational measures of affective quality and literacy-related interactions are considered separately. More than half of the correlations between the affective quality observational measures and the HOME-SF subscales are significant. By contrast, a small proportion of literacy-related observational measures are significantly correlated with the HOME-SF subscales.

Table 7.3 shows wide differences in the correlations for the ratings of mother-child book reading interactions, depending on what is being measured. The rating of Book Reading Quality (which encompasses Mother's Reading Intonation and Animation, Reading Fluency, and Comfort Level in the book reading task) is quite consistently related to the HOME-SF subscales (with only the correlations with the HOME-SF Harsh Discipline subscale not significant). By contrast, the measures of discrete behaviors (Total Number of Utterances, Percentage of Immediate Utterances, Number of Nonimmediate Utterances, and Number of Discussion Topics) show low and

Table 7.4

Summary Information: Correlations of Observational and Interview-Based Measures of Parenting

Observational Measure	Interview-Based Measure	Range of Correlations (in absolute terms)	Proportion of Significant Correlations	Variable Showing Highest Correlation
Affective Quality	HOME-SF	.02- .27	15/25 (.60)	Mother's Supportive Presence with HOME Total; Quality of Relationship with HOME Total
	Maternal Report Scales	.01-.25	2/15 (.13)	Mother's Supportive Presence (negatively) with Control
	Time Use	.02-.13	3/10 (.30)	Mother's Harsh Treatment (negatively) with Overall Parenting Time; Mother's Harsh Treatment (negatively) with Parenting Chore Time
Literacy: Book Reading Task	HOME-SF	.01-.29	4/25 (.16)	Book Reading Quality with HOME Total
	Maternal Report Scales	.01-.27	2/15 (.13)	Book Reading Quality (negatively) with Control
	Time Use	.02-.06	0/10	
Literacy: Wheels Task	HOME-SF	.01-.18	4/15 (.27)	Mother's Ease of Ideas with Cognitive Stimulation Subscale
	Maternal Report Scales	.03-.10	0/9	
	Time Use	.01- .10	0/6	

Sources: New Chance 18-month survey, brief interview accompanying observational session, and coded observational study variables.

nonsignificant correlations with the HOME-SF subscales. The discrete measures may be tapping aspects of maternal behavior, such as use of decontextualized language, that are simply not considered at all in the HOME-SF ratings. Alternately, there may be a tendency for significant correlations to occur particularly between more molar observational and interview-based measures.

3. **The Occurrence of Specific Correlations That Would Be Predicted *A Priori.*** We have distinguished between two broad domains of parenting behavior: socioemotional and cognitive. On *a priori* grounds, one would predict significant correlations between observational and interview-based measures of emotional support and, similarly, between observational and interview-based measures of cognitive stimulation.

Significant correlations occur where predicted for the observational and HOME-SF measures of the positive aspects of the mother-child relationship. That is, the observational measures of Mother's Supportive Presence, Child's Affection to Mother, and Quality of Relationship are all significantly (though not strongly) correlated with the Emotional Support subscale of the HOME-SF. However, parallel correlations are *entirely lacking* between the observational measures and the Warmth scale, the measure that relies entirely on maternal report. Indeed we note that the Warmth scale shows no significant correlations whatsoever with observational measures.

In a similar way, while one might predict that the observational measures of the positive characteristics of the mother-child relationship (Mother's Supportive Presence, Child's Affection to Mother, Quality of Relationship) would be negatively correlated with the Stress scale (also a measure dependent upon maternal report), no such significant correlations are found. Indeed, the Parenting Stress measure shows no significant correlations with any of the observational measures.

What of the several measures of disciplinary practice? Here all relationships that one might predict are confirmed. The observational measure of Mother's Harsh Treatment is indeed significantly correlated with both the HOME-SF Harsh Discipline subscale and the Maternal Control scale. We note that while significant, these correlations are quite low (-.13 between Harsh Treatment and Harsh Discipline and .14 between Harsh Treatment and Control). While related, these measures can be considered distinct, a conclusion further supported by the fact that program impacts were found for only one of these measures (Harsh Treatment). It is noteworthy that the Maternal Control measure correlates not only in the predicted manner with the observational Harsh Discipline measure, but also with several further observational measures. The pattern for the Control measure thus stands apart from that of the other two Maternal Report Parenting scales.

Are the observational measures of stimulation of literacy significantly related to the HOME-SF Cognitive Stimulation subscale, as we have predicted? Four of the eight observational measures reflecting on the mother's stimulation of her preschooler's literacy are significantly correlated with the HOME-SF Cognitive Stimulation subscale. Again, those observational measures *not* significantly correlated with the HOME-SF Cognitive Stimulation subscale concern individual verbalizations (Total Number of Utterances, Number of Nonimmediate Utterances, Percentage of Immediate Utterances, and Number of Discussion Topics). As noted above, the lack of correlation may reflect the level of the ratings (molar/molecular), or may reflect a lack of overlap in the content addressed.

4. Summary. Thus far, the empirical examination of the differing kinds of parenting measures included in the New Chance Observational Study effectively rules out the conclusion that the measures are so highly correlated that they can be considered largely overlapping measures. While the different types of parenting measures show interesting intercorrelations, the extent and magnitude of the significant relations are not great. While we can say with some certainty that the observational measures are *distinctive*, we must still ask what the particular value is of adding such measures to an evaluation of this kind, given that their collection is extremely labor-intensive and expensive and that they may be available only for a subset of families.

C. The "Value Added" of the Observational Measures of Parenting

The last step in our examination of the different parenting measures is to assess their relative effectiveness in predicting child outcomes. Do measures of parenting (from the 18-month follow-up and the observational study session at about 21 months) significantly predict child outcomes from the 42-month follow-up? Is this the case for parenting measures irrespective of the informant upon which they rely (that is, mother, interviewer, or both)? Do the observational measures provide "value added" over and above interview-based parenting measures as predictors of child outcomes?

1. A Strategy for Assessing the "Value Added" of Observational Measures. To examine these questions we selected five child outcome measures from among the longer list of child outcomes documented as part of the 42-month follow-up study of the New Chance Evaluation. We chose here to focus on total scores from the child outcome measures rather than on subscale scores, selecting specifically the total scores from the mother and teacher reports of both the Behavior Problems Index and the Positive Behavior Index. In addition, we selected the School Readiness Component of the Bracken Basic Concept Scale, a direct child assessment (see Chapter 1 for descriptions of these child outcome measures and their characteristics). We also narrowed our focus in terms of parenting behaviors considered. We sought to focus on measures of *specific parenting behaviors and inputs to the child.* As a result we set aside the Parenting Stress measure, which reflects more on maternal subjective reactions to the parenting role rather than directly on parenting behaviors and inputs; the Time Use measures, as not reflecting specifically enough on the quality of parenting behavior; and the observational variables that address child behavior rather than maternal behavior. The remaining parenting measures seek to address in differing ways the cognitive stimulation and emotional support available to the child as predictors of the selected child outcomes.[5]

To highlight differences that may relate to data collection strategies, we have distinguished among the parenting measures according to informant in these analyses. That is, we consider separately as predictors of the child outcomes the Maternal Report Parenting scales (that require only that questions be asked directly of the mother), the HOME-SF measures (that require both maternal report and ratings completed by interviewers in the home), and the observational measures (that require videotapes of mother-child interaction to be coded by outside observers). By examining the ability of each of these types of parenting measures to predict child outcomes, we will be able to shed light on whether maternal report measures alone suffice as predictors of child outcomes or whether other, more intensive data collection strategies appear warranted. It is important to note that we address this issue *only for the particular measures of each type included here*. It will be important for other studies, using different (and perhaps more exhaustive) sets of maternal report, "hybrid," and observational measures, to examine how well these measures predict child outcomes.

---

[5] In these analyses we use parenting variables from about 18 months and 21 months from baseline to predict to child outcomes about 42 months from baseline. It is not our assumption that parenting behavior stays constant across this period. Indeed, we assume that parenting behavior changes in keeping with the development of the child. Rather we are seeking evidence here that parenting at one point in time (as measured using differing approaches) is one factor in shaping later development.

Our analyses employ ordinary least squares multiple regression evaluating a series of models that each rely on different combinations of variables as predictors of our five child outcomes. We incrementally build up to an examination of whether observational measures of parenting improve prediction over and above the other types of measures. We begin with a baseline model (Model 1), which includes background characteristics of the mother and family, as well as group (experimental or control) within the evaluation. These background variables are the same ones controlled for in Chapters 4–6 in each of the impact analyses of the parenting variables. In Model 2 we add Maternal Report Parenting scales and ask whether we improve our ability to predict child outcomes. Following the same logic, Model 3 adds the HOME-SF measures to the baseline model, and Model 4 adds the observational measures to the baseline model. In each case we are able to ask whether adding a particular set of parenting variables improves our ability to predict the child outcomes beyond the consideration of background variables alone. Our next step is to isolate the specific contribution, if any, of the observational measures in predicting child outcomes. To do this, we first estimate Model 5, which includes background characteristics and all parenting variables *except* the observational variables as predictors of child outcomes. Model 6 then adds the observational measures and permits us to ask whether these measures improve our ability to predict child outcomes beyond all the other measures.

2. Summary Statistics. Tables 7.5–7.9 provide a summary of key findings of the multiple regression analyses for each of the five child outcomes. In these tables, a number of the summary statistics (specifically those that involve R^2, adjusted R^2, or difference scores for these) relate to the same underlying issue of the "proportion of variance in a child outcome explained" by a particular group of variables. This concept refers to the fact that an individual child's score on a particular child outcome is likely to vary to some extent from the group average, and it is this variation that we try to explain in regression analysis. When we include a set of variables in a model as predictors of a child outcome, we want to know how much of this variation in individual scores can be accounted for through knowledge of our chosen predictor variables. Thus, the "proportion of variance explained" indicates the amount of variation in individual scores that a set of predictor variables explains.

One way to visualize this concept is to picture a dartboard with a series of concentric circles surrounding a bull's eye. If we aim with perfect precision, we will hit the bull's eye — for example, predicting individual children's scores on the Bracken Basic Concept Scale with 100 percent accuracy. One almost never achieves this level of precision with social science variables,

Table 7.5

Summary Statistics for Analyses Examining Sets of Parenting Variables as Predictors of Mother-Reported Behavior Problems Index

Summary Statistic	Model 1 Control Variables Only	Model 2 Controls with Maternal Report Scales	Model 3 Controls with HOME-SF Scales	Model 4 Controls with Observational Measures of Parenting	Model 5 Controls with Maternal Report and HOME-SF Scales	Model 6 Controls with All Measures of Parenting
R^2	.05	.06	.11	.12	.11	.16
Adjusted R^2	.02	.02	.07	.06	.06	.08
F-statistic for model	1.56	1.45	2.56***	1.97**	2.21***	2.09***
Models being compared		2 versus 1	3 versus 1	4 versus 1	5 versus 2	6 versus 5
Difference in R^2 for above comparison		.01	.06	.07	.05	.05
Difference in adjusted R^2		.00	.05	.04	.04	.02
F-statistic for difference		1.01	5.03***	2.31**	4.52***	1.80*

Sources: New Chance baseline enrollment data, 18-month follow-up survey, coded observational study variables, and 42-month follow-up survey.

Notes: Results rely on data for the 247 observational study mothers who responded to the Behavior Problems Index at the 42-month follow-up survey. Statistical significance levels are indicated as *** < = 1 percent, ** <= 5 percent, and * <= 10 percent, which refers to the probability that the R^2 is larger than zero because of chance alone.

Table 7.6

Summary Statistics for Analyses Examining Sets of Parenting Variables as Predictors of Mother-Reported Positive Behavior Index

Summary Statistic	Model 1 Control Variables Only	Model 2 Controls with Maternal Report Scales	Model 3 Controls with HOME-SF Scales	Model 4 Controls with Observational Measures of Parenting	Model 5 Controls with Maternal Report and HOME-SF Scales	Model 6 Controls with All Measures of Parenting
R^2	.11	.14	.14	.17	.16	.21
Adjusted R^2	.08	.10	.10	.11	.11	.13
F-statistic for model	3.72***	3.74***	3.61***	2.89***	3.43***	2.80***
Models being compared		2 versus 1	3 versus 1	4 versus 1	5 versus 2	6 versus 5
Difference in R^2 for above comparison		.03	.03	.06	.05	.05
Difference in adjusted R^2		.02	.02	.03	.01	.02
F-statistic for difference		3.49**	3.06**	1.92*	2.21*	1.64

Sources: New Chance baseline enrollment data, 18-month follow-up survey, coded observational study variables, and 42-month follow-up survey.

Notes: Results rely on data for 246 observational study mothers who responded to the Positive Behavior Index at the 42-month follow-up survey. Statistical significance levels are indicated as *** <= 1 percent, ** <= 5 percent, and * <= 10 percent, which refers to the probability that the R^2 is larger than zero because of chance alone.

Table 7.7

Summary Statistics for Analyses Examining Sets of Parenting Variables as Predictors of Teacher-Reported Behavior Problems Index

Summary Statistic	Model 1 Control Variables Only	Model 2 Controls with Maternal Report Scales	Model 3 Controls with HOME-SF Scales	Model 4 Controls with Observational Measures of Parenting	Model 5 Controls with Maternal Report and HOME-SF Scales	Model 6 Controls with All Measures of Parenting
R^2	.19	.20	.20	.25	.22	.27
Adjusted R^2	.15	.15	.14	.16	.14	.15
F-statistic for model	4.14***	3.52***	3.20***	2.76***	2.89***	2.26***
Models being compared		2 versus 1	3 versus 1	4 versus 1	5 versus 2	6 versus 5
Difference in R^2 for above comparison		.01	.01	.06	.02	.05
Difference in adjusted R^2		.00	-.01	.01	-.01	.01
F-statistic for difference		1.04	.74	1.31	.81	1.18

Sources: New Chance baseline enrollment data, 18-month follow-up survey, coded observational study variables, and 42-month follow-up survey.

Notes: Results rely on data for the 148 observational study mothers and teachers who responded to the Behavior Problems Index at the 42-month follow-up survey. Statistical significance levels are indicated as *** <= 1 percent, ** <= 5 percent, and * <= 10 percent, which refers to the probability that the R^2 is larger than zero because of chance alone.

Table 7.8

Summary Statistics for Analyses Examining Sets of Parenting Variables as Predictors of Teacher-Reported Positive Behavior Index

Summary Statistic	Model 1 Control Variables Only	Model 2 Controls with Maternal Report Scales	Model 3 Controls with HOME-SF Scales	Model 4 Controls with Observational Measures of Parenting	Model 5 Controls with Maternal Report and HOME-SF Scales	Model 6 Controls with All Measures of Parenting
R^2	.19	.19	.21	.27	.22	.29
Adjusted R^2	.15	.14	.15	.18	.14	.18
F-statistic for model	4.31***	3.42***	3.48***	3.13***	2.97***	2.59***
Models being compared		2 versus 1	3 versus 1	4 versus 1	5 versus 2	6 versus 5
Difference in R^2 for above comparison		.00	.02	.08	.03	.10
Difference in adjusted R^2		-.01	.00	.03	.00	.04
F-statistic for difference		.10	1.21	1.77*	1.36	1.76*

Sources: New Chance baseline enrollment data, 18-month follow-up survey, coded observational study variables, and 42-month follow-up survey.

Notes: Results rely on data for the 152 observational study mothers and their children's teachers who responded to the Positive Behavior Index at the 42-month follow-up survey. Statistical significance levels are indicated as *** <= 1 percent, ** <= 5 percent, and * <= 10 percent, which refers to the probability that the R^2 is larger than zero because of chance alone.

Table 7.9

Summary Statistics for Analyses Examining Sets of Parenting Variables as Predictors of Scores on Bracken Basic Concept Scale

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Summary Statistic	Control Variables Only	Controls with Maternal Report Scales	Controls with HOME-SF Scales	Controls with Observational Measures of Parenting	Controls with Maternal Report and HOME-SF Scales	Controls with All Measures of Parenting
R^2	.10	.10	.13	.13	.13	.16
Adjusted R^2	.07	.06	.09	.07	.09	.08
F-statistic for model	3.26***	2.73***	3.25***	2.16***	2.85***	2.04***
Models being compared		2 versus 1	3 versus 1	4 versus 1	5 versus 2	6 versus 5
Difference in R^2 for above comparison		.00	.03	.03	.03	.03
Difference in adjusted R^2		-.01	.02	.00	.03	-.01
F-statistic for difference		.63	3.02**	1.05	3.02**	.77

Sources: New Chance baseline enrollment data, 18-month follow-up survey, coded observational study variables, and 42-month follow-up survey.

Notes: Results rely on data for the 252 observational study children who were administered the direct-assessment Bracken measure at the 42-month follow-up survey. Statistical significance levels are indicated as *** <= 1 percent, ** <= 5 percent, and * <= 10 percent, which refers to the probability that the R^2 is larger than zero because of chance alone.

however. When it comes to predicting such things as behavior and achievement, our darts tend to land much farther from the center, achieving a much lower level of prediction. The question we ask with various regression models, however, is "How much does knowledge of particular variables improve our aim?" For example, we ask "To what extent are individual scores on the Bracken explained by knowledge of family background characteristics?" "How much more of the variation in individual scores is explained, or is our metaphorical 'aim' improved, when we then add different kinds of parenting variables?" To translate from an R^2 statistic to a percentage, one simply multiplies the R^2 by 100. For example, an R^2 of .15 indicates that 15 percent of the variation in individual scores has been explained.

Tables 7.5–7.9 show:

- the R^2, or proportion of variance in the child outcome that is explained by the variables included in each model *for our particular sample*;
- the *adjusted R^2*, an estimate of the proportion of variance explained *for the population our sample is drawn from as a whole* rather than our particular sample, which *takes into account the number of independent variables considered in predicting to the outcome;* this is a conservative estimate of the proportion of variance explained and is expected to be smaller than the unadjusted R^2 ;
- the *F-statistic and its significance level*, which indicate whether the overall model being considered explains a statistically significant proportion of the variance in the child outcome; this figure is the same for the unadjusted and adjusted R^2;
- an identifier for *the particular pair of models that are being compared* in seeking to measure the further variance explained when an additional group of variables is added to a previous one in predicting the child outcome;
- *the difference in R^2* from one specified model to another, or the *additional* variance explained by inclusion of a further set of variables in the model (calculated simply by subtracting the R^2 in a previous model from that in the current model);
- *the difference in adjusted R^2*; the same statistic as above, but relying on the more conservative adjusted R^2 figure; and
- the *F-statistic and significance level for the difference in R^2*, which indicate whether the addition of a set of variables in a model beyond that in the previous model provides a significant *increase* in the proportion of variance explained in the child outcome; this figure is the same for differences in adjusted and unadjusted R^2.

Prior to carrying out the analyses summarized in these tables, we sought first to winnow the large list of possible parenting variables on conceptual as well as empirical grounds. In particular, the research teams from the University of Minnesota and Harvard University identified those variables that their previous work had identified as of particular importance conceptually and in predicting child outcomes. In addition, exploratory analyses with the full list of parenting variables were carried out to examine intercorrelations of the variables and the role of individual variables as predictors of the outcomes. The following parenting variables were included in the multiple regressions in the three groupings noted: (1) Maternal Report Parenting scales: Warmth and Control; (2) HOME-SF scales: Emotional Support, Cognitive Stimulation, and Harsh Discipline; and (3) observational measures: Mother's Harsh Treatment, Mother's Confidence, Mother's Intrusiveness, Mother's Supportive Presence, Percentage of Immediate Utterances, Book Reading Quality, Mother's Ease of Ideas, and Wheeled Objects Named as a Proportion of Maternal Elicitations.

We turn now to a summary of our findings for two key questions: whether each set of parenting variables, considered separately, improves prediction of child outcomes beyond the consideration of control variables alone; and whether adding the observational variables improves prediction of child outcomes when all other parenting variables under consideration here have already been taken into account.

How well does each set of parenting variables (Maternal Report Parenting scales, HOME-SF, and observational) predict child outcomes? Does each set, considered separately, improve prediction of child outcomes beyond the background characteristics alone? By examining the F-statistic for Model 1 of Tables 7.5–7.7, we can see that the background variables alone explain a significant proportion of the variance for four of the five child outcome variables. Only for the mother's report of the Behavior Problems Index does taking into account the eight background or control variables fail to explain a significant proportion of the variance.

When only the Maternal Report Parenting scales are added to the background variables (Model 2), it is generally not the case that we see a significant improvement in predicting each of the child outcomes. The F-statistic for the difference in the proportion of variance explained for Model 2 versus Model 1 is significant for only *one* child outcome, the mother's report of the Behavior Problems Index. By contrast, in parallel analyses, when only the HOME-SF subscales are added to the background variables (Model 3), the F-statistic for the difference (Model 3 versus Model 1) indicates a significant increase in the proportion of variance explained for *three* of the child outcomes: mother's report of both problem and positive child behavior as well as the assessment of cognitive development. Similarly, in Model 4, adding the observational variables to the background measures significantly increases the ability to predict *three* child outcomes: mother's report of behavior problems and positive behaviors, and teacher's report of positive behaviors. In general, the HOME-SF and observational measures of parenting appear to be stronger predictors of the child outcomes than the Maternal Report Parenting scales considered here.

Do the HOME-SF subscales improve prediction of child outcomes over and above the Maternal Report Parenting scales? Do observational measures improve prediction to a still greater degree? Our next questions focus on the specific contribution, if any, of using measures from different informants to predict child outcomes. First, when background variables and the Maternal Report Parenting scales have already been entered as predictors, does adding the HOME-SF variables result in a significant increment in the proportion of variance explained in the child outcomes? Second, and of central importance in the present context, does adding observational variables to the full complement of background and other parenting variables increase our ability to predict the child outcomes?

In the tables, results for Model 5 (specifically the significance level of the F-statistic for difference in proportion of variance explained in Model 5 versus Model 2) indicate that for *three* of the child outcomes of interest (mother's report of both behavior problems and positive behaviors, as well as the assessment of cognitive development), adding the HOME-SF variables to the background measures and the Maternal Report Parenting scales significantly increases our ability to predict the outcome. Thus, the HOME-SF measures, which rely upon a combination of maternal report and interviewer ratings, improve our ability to explain variation in these three child outcomes beyond what simple maternal report permits in combination with background variables.

What about the further addition of the observational variables? When background characteristics, the Maternal Report Parenting scales, and the HOME-SF subscales have been taken into account to predict child outcomes, what is the "value added," if any, of the inclusion of observational measures? The F-statistic for the difference in proportion of variance explained in Model 6 versus Model 5, for each child outcome, indicates that the observational variables do indeed provide "value added" in this sense for two of the child outcomes: mother's report of the behavior problems and teacher's report of positive behaviors.

3. Summary. In general, it appears that inclusion of an outside perspective, that of an interviewer in the hybrid HOME-SF subscales, or complete reliance on an outside observer in the observational measures, strengthens our ability to predict child outcomes beyond measures relying entirely on maternal report. Indeed, our parenting scales based entirely on maternal report do not provide a strong basis for predicting child outcomes. It is important again to note the caveat that this conclusion pertains only to the two Maternal Report Parenting scales included here and to our particular sample. The conclusion might be different for a more comprehensive battery of maternal report measures or for a different or more heterogeneous sample of families. It is particularly important to note that with all other parenting measures considered here already included in the model, adding the observational measures improves prediction of two of the child outcomes. Within our sample, as for samples studied by Berlin and colleagues (1995, 1996), observational measures appear to provide a perspective that increases our ability to explain variation in some child outcomes.

D. Conclusion

We see three distinct contributions that the observational variables make to an evaluation like that of the New Chance Program. First, the observational measures provide information that could not be obtained from the interview-based measures of parenting. For example, it is only through the observational measures in the present study that we obtain a picture of the quality (as opposed to quantity) of the cognitive stimulation available to children. Indeed, it is only through the observational measures that we see any indication of a program impact in the domain of cognitive stimulation provided to children within this evaluation. Observational measures may thus detect differences that the interview-based measures cannot.

Second, where there is concordance between findings from the observational and interview-based measures of parenting, because of the rigor with which the observational measures were obtained, our confidence in the overall pattern of findings is increased. Only the observational measures were completed by coders blind as to research group. The finding of a difference in observed affective quality of mother-child interaction (less harsh treatment in the experimental group) increases our confidence that children in the experimental group experienced a different socioemotional climate, a finding suggested also by differences on the HOME-SF and Maternal Warmth scale. In the absence of the observational study findings, the findings derived from interviews would be open to greater speculation about the contribution of response biases by mothers and/or interviewers.

Third, we have seen that the observational measures are significant predictors of several important child outcomes. Further, our analyses indicate that we *improve* our ability to predict some child outcomes when observational variables are added to Maternal Report Parenting scales and HOME-SF variables as well as family background measures, a stringent examination of the ability of the observational measures to predict child outcomes. In recent work, Berlin and colleagues (1995, 1996) found that direct observation of mother-child interaction in a teaching task more consistently predicted child outcomes than did HOME Inventory subscales in a sample of African American families. Thus, the evidence to date from two research teams supports the conclusion that inclusion of observational variables improves our ability to understand how parenting contributes to children's development.

Chapter 8

Parenting in a Broader Context: An Examination of the Multiple Influences on Child Outcomes

New Chance Observational Study Research Team

This chapter seeks to address a paradox inherent in the results presented to this point. We have documented positive impacts of the New Chance Program on parenting behavior and links between parenting behavior and later child outcomes within the observational study sample. Yet we see an absence of positive program impacts on the development of children in our sample, and indeed some evidence of negative program impacts. This pattern suggests the need to look at factors other than parenting as playing a role in shaping children's development. In this chapter we look at the importance of a broader set of variables in predicting children's development within the observational study sample. In addition to background characteristics and parenting behavior, we consider mother's subjective well-being, mother's human capital (employability and employment), and the larger social context of the family. Results suggest a pattern in which positive effects of supportive and stimulating parenting behavior combine with unfavorable effects of maternal psychological distress and stress in the broader social context in influencing several child outcomes. The findings point to the necessity of addressing not only parent-child relations, but also the context of the family, if we are to succeed in bringing about positive effects on the development of children in families of teenage mothers on welfare.

The findings summarized to this point for the New Chance Observational Study present us with a paradox. The results noted in Chapters 4–6 confirm that New Chance had positive impacts on parenting behavior. Impacts are found not only on parenting measures that rely on maternal report or interviewer ratings but also on the rigorous observational measures, thus increasing our confidence in the overall pattern. Further, we noted in Chapter 7 that measures of parenting are significant predictors of child outcomes within the observational study sample. Parenting behavior that is less harsh and more emotionally supportive, and that provides a higher level of cognitive stimulation, is associated with better developmental outcomes in the children

---

NOTE: This chapter was conceptualized and developed collaboratively by the full team, whose members (in alphabetical order) include: Hans Bos, Jeanne De Temple, M. Robin Dion, Byron Egeland, Carolyn Eldred, Robert Granger, Donna Ruane Morrison, John Ogawa, Catherine Snow, Nancy Weinfield, and Martha Zaslow.

about 42 months after the start of the evaluation. Thus, New Chance had positive program impacts on parenting, and parenting behavior within this sample mattered to children's later development.

Such a pattern might lead us to expect that there would be positive program impacts on measures of children's development: that children within the New Chance experimental group would show better developmental outcomes at the 42-month follow-up. Yet this is not the case. As summarized in Chapter 2, on most child outcome measures there is no evidence of a difference in the development of children in the experimental and control groups within the observational study sample. For example, we found no program impact on the direct assessment of children's cognitive development. Teachers' reports of children's school adjustment and academic progress, for those children already participating in school settings, did not differ for children in the New Chance experimental and control groups. Further, some maternal report measures of the children's development reflected *negative* program impacts. Mothers in the New Chance experimental group perceived their children's social behavior less favorably. Within the observational study sample, mothers in the experimental group described their children as showing significantly fewer positive social behaviors and as differing specifically in the important areas of social competence and compliance with adult requests.

How can we explain this absence of positive program impacts on children's behavior, and indeed the indications of negative program impacts on their social behavior, when we have seen positive program impacts on parenting and a linkage between parenting behavior and child outcomes? Perhaps we need to broaden our scope beyond background characteristics and parenting behavior in considering the factors that help to shape outcomes for children in the New Chance sample. While developmental psychologists had long accorded parenting behavior a special status as the major or sole contributor to children's development, growing evidence has challenged this view. Rather, it appears that a wider range of relationships and contexts must be taken into account.

Bronfenbrenner (1979, 1986), for example, contends that if we wish to understand children's development we must consider not only relationships within the family but also the larger *ecology* within which the family functions:

> Although the family is the principal context in which human development takes place, it is but one of several settings in which developmental processes can and do occur. . . . The psychological development of children in the family is affected not only by what happens in the other environments in which children spend their time but also by what occurs in the other settings in which their parents live their lives. (1986, p. 723)

Bronfenbrenner presents evidence for the importance to children's development of such further environments that children experience directly as child care, school, and neighborhood. The parents' workplace is an example of an environment with which children may have limited contact, yet which may have important implications for their development. For example, mothers may emphasize in their socialization practices at home, behaviors (for example, self-direction) for which they are rewarded on the job, and work-related stress may affect relationships at home.

Other researchers taking an "ecological perspective" have focused specifically on economic factors, especially economic stress (McLoyd, 1990), and on the degree to which the mother is socially isolated or supported (Belsky, 1984), as having implications for children's development.

New Chance targeted families of poor young mothers with limited education. As we have noted in Chapter 1, many of the young mothers in the New Chance Evaluation were facing serious stressors in terms of their economic and housing circumstances and their relationships with parents and partners. Perhaps the larger context has unusual importance for mothers facing multiple stressors and plays a key role in the development of their children as well.

Further, we have seen indications that the New Chance Program had effects on variables that may be seen as part of the broader context of children's development. For example, Chapter 2 documented positive program impacts within the observational study sample on maternal educational attainment, an important component of the family's economic situation. Yet we have also seen that there were unintended program impacts on such features of the broader context as living arrangements and frequency of changes in residence. Thus, it seems particularly important that we consider extrafamilial as well as intrafamilial factors as contributors to children's development within the New Chance Observational Study sample.

In this chapter we broaden our consideration of possible contributors to children's development. Chapter 7 focused intensively on parenting variables and family background characteristics; in this chapter we consider additional variables reflective of the mother's psychological well-being; the mother's employability, actual employment, and earnings; and contexts beyond the parent-child relationship (such as living situation and child care participation) that the mothers and children have experienced.

I. Analysis Strategy

The strategy of analysis used here parallels that used in Chapter 7 in that we again examine the relative contribution of different sets of variables in predicting selected child outcomes. However, here we encompass a broader set of predictor variables and organize these variables into somewhat different groupings. We turn now to a description of the child outcome variables and the predictor variables selected for these analyses and the strategy of analysis.

A. Child Outcomes Examined

For the present analyses we selected five child outcomes from among those collected as part of the 42-month follow-up of the New Chance Evaluation. We focus here on the same five outcomes examined in Chapter 7: the Behavior Problems Index and the Positive Behavior Index as reported on by both mother and teacher, and the Bracken Basic Concept Scale School Readiness Component, a direct assessment of child cognitive development. To limit the number of analyses, we again chose to focus only on the *total* scores of the Behavior Problems Index and the Positive Behavior Index. A higher score on the Behavior Problems Index indicates a perception (by mother or teacher) of more behavior problems, whereas a higher score on the Positive Behavior Index indicates a perception of more positive social behavior. A higher score on the Bracken School Readiness Component indicates greater mastery of concepts directly relevant to school (for example, letter identification, numbers, shapes, and comparisons). The outcomes chosen permit us to examine the children's cognitive development, behavior problems, and positive social development. As noted in the 42-month report of the New Chance Evaluation (Quint, Bos, and Polit, 1997), there has been a tendency in social science to focus on negative social outcomes, particularly among children in poverty. We consider it essential to examine positive social outcomes as well.

The examination of two parallel child outcome measures reported on by both mother and teacher (the Behavior Problems Index and the Positive Behavior Index) permits us to explore the question of whether similar or different predictors help to explain children's scores on these outcomes when reported on by these two different informants. We have already seen some indication that mothers and teachers differ in their perceptions of the children in the New Chance Evaluation (see Chapters 1, 2, and 7). For example, within the observational study sample, mothers, but not teachers, saw children in the experimental group as less socially competent and less compliant. We wondered, then, whether the mother's and teacher's reports of the same child outcome would be best predicted by the same or different predictor variables.[1]

B. Sets of Predictor Variables

Five sets of variables will be considered as predictors of these child outcomes: background or control, parenting, mother's subjective well-being, mother's human capital (employability, employment and earnings), and larger social context. Apart from the background or control variables, each set of variables was chosen on the dual grounds that previous research substantiates the importance of the set of variables in helping to shape child outcomes and that this was an area of the mother's or child's life that the New Chance Program had identified as an important area to target in the intervention.

1. Control Variables (from Baseline Data). We used the same set of background or control variables here as were used in the analyses of different kinds of parenting measures as predictors of child outcomes in Chapter 7. These variables are (1) race/ethnicity, (2) focal child's gender, (3) focal child's age, (4) number of children the mother had given birth to, (5) mother's TABE score, (6) the Philadelphia site, (7) the Portland site, and (8) New Chance experimental/control group status. These variables are all "baseline characteristics," that is, derived from data collected prior to or at the time of mother's random assignment to the experimental or control group within the evaluation, and thus unaffected by the intervention.

2. Parenting Variables (from the Period of the Evaluation's First Follow-Up).[2] Parenting variables were considered in these analyses because of their acknowledged importance in helping to shape children's development (Bornstein, 1995; Maccoby and Martin, 1983) and because New Chance sought to improve mothers' interactions with their children as well as mothers' provision of stimulation through the home environment. Indeed, our analyses were intended, in part, to examine the *relative* importance of parenting variables when other sets of variables were simultaneously considered as predictors of child outcomes.

---

[1] It is important to note that the sample size for the analyses based on mother's and teacher's reports differs. The teacher's report measures are based on the smaller sample of children already in school or a formal child care arrangement whose mothers consented for the child's teacher to be contacted and whose teachers returned the Teacher Questionnaire. Thus, the analyses of the same outcomes based on teacher's and mother's reports are not based on identical samples. We can examine the pattern of predictors for these two kinds of child outcomes and note similarities and differences. However, we must include the difference of samples as one possible basis for interpreting any difference in which predictor variables are significant.

[2] The period of the study's first follow-up includes both the 18-month follow-up survey carried out in the full evaluation sample, and the observational session, carried out an average of 21 months after random assignment.

The analyses reported on at the end of Chapter 7 gave separate consideration to three different types of parenting variables: Maternal Report Parenting scales, HOME-SF subscales, and observational measures. Here we consider parenting measures in a single set, irrespective of informant. In the present analyses we build on the Chapter 7 analyses by choosing the individual parenting measures that emerged as the strongest measures in those analyses. Parenting variables were selected on the grounds that each one had had a significant coefficient in predicting at least one of the five child outcomes considered in the Chapter 7 analyses. Eight parenting variables were selected on this basis: (1) the Maternal Report Warmth scale, (2) the HOME-SF Emotional Support subscale, (3) the HOME-SF Cognitive Stimulation subscale, (4) the HOME-SF Harsh Discipline subscale, (5) the observational measure of Mother's Harsh Treatment, (6) the observational measure of Mother's Confidence, (7) the observational measure of Percentage of Immediate Utterances, and (8) the observational measure of Wheeled Objects Named as a Proportion of Maternal Elicitations. All of the parenting measures were derived from the period of the first follow-up within the New Chance Evaluation, either the 18-month follow-up survey or observational session carried out about 21 months after random assignment.

3. Mother's Subjective Well-Being Variables (from the Period of the Evaluation's First Follow-Up). Measures of maternal subjective well-being fulfill our dual requirements for selection of further predictor variables in the present analyses. That is, previous research confirms the importance of such variables as predictors of child outcomes, and improving maternal emotional well-being was a goal of the New Chance Intervention.

A review of the evidence on the implications for children of maternal depression (Downey and Coyne, 1990), for example, found that children of depressed mothers have higher levels of both externalizing (aggressive) and internalizing (anxious, depressed) behavior problems. Other research documents the importance of such further aspects of subjective well-being as mothers' sense of mastery or control over events in their lives (Stevens, 1988) and sense of stress in the parenting role (Abidin, 1983). Furthermore, the New Chance Program was "specifically structured to foster positive emotional growth among the participants. . . . The goal of improving the emotional well-being of participants was consistent with the considerable evidence indicating that high levels of depression and stress constitute important barriers to effective functioning in adult roles. . . ." (Quint et al., 1994, p. 158).

For the current analyses we selected measures from the period of the first follow-up (the 18-month survey and the additional data collection visit for the observational study sample carried out about 21 months after random assignment) that reflected on mothers' *internal* sense of well-being. We thus distinguished between measures reflecting on mothers' internal or subjective sense of well-being (for example, control, satisfaction, depression, and stress) and measures reflecting sources of difficulty and support in the broader social context (for example, the Difficult Life Circumstances and social support measures), which we included among measures of the larger social context. We selected only measures from the first follow-up (rather than 42-month measures of depression, stress, and so on), because of our concern that measures from the later period could be seen not only as *shaping* child outcomes, but also as *responding* to the well-being or problems with the child (the problem of "endogeneity").

Our measures of maternal subjective well-being included the following scales: (1) Mastery (Pearlin et al., 1981), a 7-item scale measuring mothers' sense of self-efficacy; (2) Parenting Stress (see Quint et al., 1994), an 8-item scale measuring mothers' subjective sense of aggravation and stress in their interactions with their children; (3) Center for Epidemiological Studies Depression (CES-D) Scale (Radloff, 1977), a 20-item scale measuring risk of clinical depression; and (4) Life Satisfaction, a 9-item scale measuring overall life satisfaction (see Chapter 2).

4. Mother's Human Capital Variables (from the 42-Month Follow-Up). We included a set of measures reflecting the mother's employability, actual employment, welfare receipt, and earnings, again because these were important priorities of the New Chance Program, and at the same time have been found in previous research to be important to children's developmental outcomes. New Chance sought in its first phase to help participating mothers complete the GED. But as we noted in Chapter 1, the New Chance Program aimed beyond completion of the GED to help mothers, in phase 2 of the program, to gain specific job skills, participate in apprenticeships, and ultimately obtain employment.

Previous research has shown that measures of mothers' employability, employment, and earnings are among the strongest predictors of children's development. For example, parents' educational attainment is closely associated with the level of schooling that children complete (Hauser and Mossel, 1985; Sewell and Hauser, 1976; Bowles and Gintis, 1976). In a recent study following the children born to teenage mothers into adulthood, the teenage mothers' educational attainment was found to predict the grown children's functional literacy in young adulthood (Baydar, Brooks-Gunn, and Furstenberg, 1993). Measures of family economic self-sufficiency, such as total earnings and income, whether or not the family receives welfare and over what time period, have all been found to predict child outcomes (Duncan, Brooks-Gunn, and Klebanov, 1994; Moore et al., 1995).

In the current analyses we included five measures of the mother's human capital and family economic self-sufficiency, all taken from the 42-month follow-up: (1) whether the mother had received a high school diploma or a General Equivalency Diploma (GED), (2) whether the mother had obtained a trade license, (3) whether the family was receiving Aid to Families with Dependent Children (AFDC), (4) whether the mother was employed, and (5) mother's total earnings across the full 42-month follow-up period.

We selected 42-month outcomes for this set of predictor variables to permit as much time as possible for the young mothers to make progress in these areas. We realize that some variables measuring mother's human capital at the 42-month follow-up are potentially endogenous, that is, potentially affected by child behavior and development. For example, it is possible that a mother's ability to maintain employment is affected by her child's behavior in school. However, since most of the variation in these outcomes is probably determined exogenously (that is, by factors other than child outcomes), we felt that it was appropriate to use 42-month outcomes as explanatory variables in these analyses.

5. Larger Social Context Variables (primarily from the 42-Month Follow-Up). Our final set of predictor variables sought to capture the nature of the social context (beyond the mother-child dyad) that mothers as well as children in the sample experienced. These variables reflect on the *living circumstances of the family* (number of moves, whether or not the mother resided with a partner or husband, number of children in household); *the mother's perception of serious difficulties as well as social support* in this broader social context; and *the child's experience of nonmaternal care* (participation in day care or preschool). Bronfenbrenner's work amply documents the role of such extrafamilial factors in children's development (1979, 1986). Further, the New Chance Program sought changes in mothers' and children's extrafamilial experiences through providing child care as part of the program and through seeking to enhance mothers' sense of social support.

We included eight variables in our consideration of the broader social context of the mothers and children in our sample: (1) a Difficult Life Circumstances scale (an adaptation of the measure developed by Barnard, 1988) documenting the mothers' experience of such stressful circumstances as having the electricity or phone cut off; being robbed, mugged, or attacked; and having a close personal relationship with someone who had died or been killed; (2) the number of times the family had moved since enrolling in the evaluation; (3) whether or not the mother was living with a partner or boyfriend; (4) the number of children living in the household; (5) the mother indicating that she had no sources of social support; (6) the mother's rating of satisfaction with her social support; (7) the number of months the focal child had been in day care or preschool; and (8) whether or not the focal child had been in child care before his or her first birthday.

With the exception of the measures of the focal child's day care participation and residence with a partner, all of these measures were taken from the 42-month follow-up. We sought to examine the implications of the *current* life circumstances of mother and child. As for the human capital development variables, we assume variation in these circumstances to be largely exogenous, that is, unaffected by differences in child outcomes. Information on history of child care participation was not collected during the 42-month follow-up, and thus we use data from the earlier survey for these variables. We note also that measures of child care quality and stability were not available for individual children in the New Chance Evaluation.[3] Thus, the aspects of child care considered here are the extent and timing of participation, rather than the nature of the child's experience in care.

---

[3] A special study of child care quality was conducted as part of the New Chance Evaluation (see Quint et al., 1994; Fink, 1995). This study involved carrying out direct observations of child care quality in selected sites with on-site child care, using the Early Childhood Environment Rating Scale and the Infant and Toddler Environment Rating Scale. However these measures assessed classroom quality rather than the experiences of individual children. The quality of child care experienced by children in the control group was not examined in this special study. Further, several of the sites from which subjects were drawn for the New Chance Observational Study were not included in the special study of on-site New Chance child care, and indeed several observational study sites did not offer on-site child care.

C. Analytic Approach

We used ordinary least squares multiple regression to examine prediction to each of the five selected child outcomes. In predicting each child outcome, we started with a baseline model (Model 1), which included only the background characteristics or control variables as predictors (labeled Control Variables in the tables). Each subsequent model *added* a block of variables cumulatively to the prior model. Model 2 added the block of variables reflecting parenting behavior (labeled Parenting); Model 3 then added variables reflecting maternal subjective well-being (labeled Mother's Subjective Well-Being); Model 4 added the block of variables concerning mothers' education, employment, and earnings (labeled Mother's Human Capital); and, finally, Model 5 added the block of variables focusing on the broader context experienced by the mother and child (labeled Larger Social Context).

Tables 8.1-8.5 present results for the prediction to the five separate child outcomes. As in Chapter 7, the summary statistics in these tables provide information as to how much of the variation in the particular child outcome is explained by knowledge of all of the variables considered in the particular model and, further, how much *additional* variation is explained by the addition of the new set of variables considered in a particular model. By "variation in the particular child outcome" we mean that individual children's scores will vary, with some children doing better than the group average and some children doing worse. When we examine the "proportion of variance explained" we are asking how well we can predict children's scores through knowledge of a set of variables that may help determine if a child does relatively better or worse. For example, when we know only families' background characteristics, can we do statistically better than chance in predicting children's varying scores? When we know families' scores on both background characteristics and parenting measures? In each instance, what proportion of the variance in children's scores can we predict through knowledge of these variables? When we add a new set of variables in our attempt to predict a child outcome, how much have we improved our ability to predict?

The F-statistic for the R^2 and adjusted R^2 is the same, and indicates whether we have explained a statistically significant proportion of the variance in the child outcome through inclusion of the variables in a model. The difference in R^2 (and adjusted R^2) figures indicate how much more variance we have explained by adding a set of variables to the previous model; it is calculated simply by subtracting the relevant statistic for the first (smaller) model from that of the second (larger) model. The F-statistic for the difference in R^2 (and adjusted R^2) tells us whether we have significantly increased the proportion of variance in the child outcome that we have explained by adding a new set of variables. As in Chapter 7, we interpret results based on both the R^2 and the more conservative adjusted R^2. While Tables 8.1-8.5 present summary statistics from the multiple regression analyses, Appendix Tables C.1-C.5 provide the full set of results from each analysis, so that the coefficients for individual variables can be examined.

Table 8.1

Summary Statistics for Analyses Examining Sets of Variables as Predictors of Mother-Reported Behavior Problems Index

Summary Statistic	Model 1 Control Variables Only	Model 2 (Adds Parenting)	Model 3 (Adds Mother's Subjective Well-Being)	Model 4 (Adds Mother's Human Capital)	Model 5 (Adds Larger Social Context)
R^2	.05	.15	.23	.25	.35
Adjusted R^2	.02	.09	.16	.17	.25
F-statistic for model	1.56	2.59***	3.42***	2.96***	3.55***
Models being compared		2 versus 1	3 versus 2	4 versus 3	5 versus 4
Difference in R^2 for above comparison		.10	.08	.02	.10
Difference in adjusted R^2		.07	.07	.01	.08
F-statistic for difference		3.49***	4.60***	1.10	4.30***

Sources: New Chance baseline enrollment data, 18-month follow-up survey, coded observational study variables, and 42-month follow-up survey.

Notes: Results rely on data for the 247 observational study mothers who responded to the Behavior Problems Index at the 42-month follow-up survey. Statistical significance levels are indicated as *** <= 1 percent, ** <= 5 percent, and * <= 10 percent, which refers to the probability that the R^2 is larger than zero because of chance alone.

Table 8.2

Summary Statistics for Analyses Examining Sets of Variables as Predictors of Mother-Reported Positive Behavior Index

Summary Statistic	Model 1 Control Variables Only	Model 2 (Adds Parenting)	Model 3 (Adds Mother's Subjective Well-Being)	Model 4 (Adds Mother's Human Capital)	Model 5 (Adds Larger Social Context)
R^2	.11	.20	.25	.27	.34
Adjusted R^2	.08	.15	.19	.19	.24
F-statistic for model	3.72***	3.64***	3.86***	3.31***	3.31***
Models being compared		2 versus 1	3 versus 2	4 versus 3	5 versus 4
Difference in R^2 for above comparison		.09	.05	.02	.07
Difference in adjusted R^2		.07	.04	.00	.05
F-statistic for difference		3.28***	2.77**	1.09	2.69***

Sources: New Chance baseline enrollment data, 18-month follow-up survey, coded observational study variables, and 42-month follow-up survey.

Notes: Results rely on data for the 246 observational study mothers who responded to the Positive Behavior Index at the 42 month follow-up survey. Statistical significance levels are indicated as *** <= 1 percent, ** <= 5 percent, and * <= 10 percent, which refers to the probability that the R^2 is larger than zero because of chance alone.

BEST COPY AVAILABLE

Table 8.3

Summary Statistics for Analyses Examining Sets of Variables as Predictors of Teacher-Reported Behavior Problems Index

	Model 1	Model 2	Model 3	Model 4	Model 5
Summary Statistic	Control Variables Only	(Adds Parenting)	(Adds Mother's Subjective Well-Being)	(Adds Mother's Human Capital)	(Adds Larger Social Context)
R^2	.19	.26	.28	.33	.38
Adjusted R^2	.15	.17	.17	.19	.21
F-statistic for model	4.14***	2.91***	2.53***	2.38***	2.17***
Models being compared		2 versus 1	3 versus 2	4 versus 3	5 versus 4
Difference in R^2 for above comparison		.07	.02	.05	.05
Difference in adjusted R^2		.02	.00	.02	.02
F-statistic for difference		1.55	.58	1.54	1.34

Sources: New Chance baseline enrollment data, 18-month follow-up survey, coded observational study variables, and 42-month follow-up survey.

Notes: Results rely on data for the 148 observational study mothers and teachers who responded to the Behavior Problems Index at the 42-month follow-up survey. Statistical significance levels are indicated as *** <= 1 percent, ** < =5 percent, and * < =10 percent, which refers to the probability that the R^2 is larger than zero because of chance alone.

Table 8.4

Summary Statistics for Analyses Examining Sets of Variables as Predictors of Teacher-Reported Positive Behavior Index

Summary Statistic	Model 1 Control Variables Only	Model 2 (Adds Parenting)	Model 3 (Adds Mother's Subjective Well-Being)	Model 4 (Adds Mother's Human Capital)	Model 5 (Adds Larger Social Context)
R^2	.19	.28	.28	.31	.39
Adjusted R^2	.15	.19	.17	.18	.22
F-statistic for model	4.31***	3.27***	2.54***	2.30***	2.28***
Models being compared		2 versus 1	3 versus 2	4 versus 3	5 versus 4
Difference in R^2 for above comparison		.09	.00	.03	.08
Difference in adjusted R^2		.04	-.02	.01	.04
F-statistic for difference		1.98*	.01	1.23	1.84*

Sources: New Chance baseline enrollment data, 18-month follow-up survey, coded observational study variables, and 42-month follow-up survey.

Notes: Results rely on data for the 152 observational study mothers and teachers who responded to the Positive Behavior Index at the 42-month follow-up survey. Statistical significance levels are indicated as *** <= 1 percent, ** <= 5 percent, and * <= 10 percent, which refers to the probability that the R^2 is larger than zero because of chance alone.

Table 8.5

Summary Statistics for Analyses Examining Sets of Variables as Predictors of Scores on Bracken Basic Concept Scale

Summary Statistic	Model 1 Control Variables Only	Model 2 (Adds Parenting)	Model 3 (Adds Mother's Subjective Well-Being)	Model 4 (Adds Mother's Human Capital)	Model 5 (Adds Larger Social Context)
R^2	.10	.14	.15	.16	.19
Adjusted R^2	.07	.08	.08	.06	.07
F-statistic for model	3.26***	2.43***	2.09***	1.70**	1.58**
Models being compared		2 versus 1	3 versus 2	4 versus 3	5 versus 4
Difference in R^2 for above comparison		.04	.01	.01	.03
Difference in adjusted R^2		.01	.00	-.02	.01
F-statistic for difference		1.54	.89	.26	1.17

Sources: New Chance baseline enrollment data, 18-month follow-up survey, coded observational study variables, and 42-month follow-up survey.

Notes: Results rely on data for the 252 observational study children who were administered the direct- assessment Bracken measure at the 42-month follow-up survey. Statistical significance levels are indicated as *** <= 1 percent, ** <= 5 percent, and * < =10 percent, which refers to the probability that the R^2 is larger than zero because of chance alone.

D. Specific Questions to Be Examined

In examining the results of these analyses we ask the following questions:

- Does more supportive and stimulating parenting behavior predict more positive child outcomes? Does adding parenting variables significantly improve prediction of child outcomes?
- Does greater maternal subjective distress (as manifested in higher depression, greater parenting stress, lower mastery scores, or less life satisfaction) predict less positive child outcomes? Does adding these variables improve prediction of child outcomes?
- Does greater maternal human capital predict more positive child outcomes? Does adding these variables improve prediction of child outcomes?
- Does stress in the larger social context predict less positive child outcomes? Does consideration of these variables improve prediction of child outcomes?
- Do the same sets — or different sets — of variables serve as significant predictors of the measures of social behavior as reported on by mothers and by teachers?
- Does an ecological perspective, which takes into account the larger social context in which parenting takes place, help explain our paradox: that is, that more supportive and stimulating parenting behavior does not result in positive program effects on child outcomes at 42 months?

We present results for the two mother-reported measures of the child's social behavior, for the two teacher-reported measures of the child's social behavior, and, finally, for the assessment of the child's cognitive development. We then examine the findings across the child outcomes in light of our specific questions.

II. Predicting Child Outcomes from the Broader Set of Variables

A. Findings for Mother-Reported Behavior Problems and Positive Behaviors

1. Behavior Problems Index. Table 8.1 presents a summary of the results for the mother's report of the Behavior Problems Index. We see that family background characteristics alone (Model 1) do not explain a significant proportion of the variance in the Behavior Problems Index as reported on by the mother. However, by adding the parenting variables (Model 2), we have explained an additional 10 percent of the variance in children's scores (or 7 percent according to the change in the more conservative adjusted R^2), which is a statistically significant increase in the proportion of variance explained. Examination of the individual parenting measures in Model 2 (see Appendix Table C.1) reveals that greater harshness in discipline, as measured by both the HOME-SF and the observational Harsh Treatment measure, predicts a greater number of behavior problems in the child. In addition, mothers whose children had higher Behavior Problems Index total scores were less confident in interacting with their children during the observational session.

Adding variables reflecting the mother's subjective well-being in Model 3 results in an increase in the proportion of variance explained of 8 percent (7 percent according to the change in adjusted R^2). These figures reflect a statistically significant increase in the proportion of variance in the child outcome explained. In particular, greater Parenting Stress at the time of the first follow-up was predictive of higher scores on the Behavior Problems Index at 42 months.

In Model 4, adding the measures of the mother's human capital does not significantly improve our ability to predict this child outcome. However, by adding the block of variables addressing the larger social context in Model 5, we can explain 10 percent more of the variance in the Behavior Problems Index (8 percent according to the change in adjusted R^2). Individual measures that were associated with higher scores on the Behavior Problems Index were the Difficult Life Circumstances scale and, unexpectedly, the measure of duration of participation in day care. The more difficult life circumstances the mother was facing, and the more months the child had participated in day care, the greater the maternal report of behavior problems.

Taking into account all of the variables included in the final model accounts for approximately one-third of the variance in this child outcome (and adjusted R^2 =.25). By either estimate, these variables explain a substantial proportion of the variation in children's scores on the Behavior Problems Index.

2. **Positive Behavior Index.** Table 8.2 presents summary statistics for the analyses predicting the Positive Behavior Index according to mother's report. Findings are broadly similar to those for the Behavior Problems Index except that the initial block of background variables does, in this instance, explain a significant proportion of the variance (R^2 = 11 percent, adjusted R^2 = 8 percent) in the outcome. Again, adding parenting variables, subjective well-being variables, and variables addressing the larger social context significantly improves our ability to predict children's scores, whereas adding the human capital development variables does not. When we examine the particular measures significantly associated with mother's report of positive child social behavior (see Appendix Table C.2), however, we see some interesting differences from those reported above for the Behavior Problems Index.

Background characteristics that were significantly predictive of mothers' report of children's positive behaviors in Model 1 (see Appendix Table C.2) included gender (girls received higher ratings) and race/ethnicity (black mothers gave higher ratings). Mothers with more than one child at baseline reported lower scores on the Positive Behavior Index. In addition, the coefficient for experimental/control group membership of the family is significant, reflecting the program impact on this measure that we reported in Chapter 2, with mothers in the experimental group giving significantly lower ratings to their children on the Positive Behavior Index. Subsequent models control for all background characteristics, including group within the evaluation.

In Model 2 the specific parenting variables that were significantly associated with this child outcome included, as was the case for the mother's report of the child's behavior problems, the HOME-SF Harsh Discipline subscale and the observational Mother's Confidence measure. However, here we also see that the Maternal Warmth scale serves as a significant predictor. The addition of the parenting variables explains a further 9 percent of the variance in children's scores (7 percent according to adjusted R^2), a statistically significant increase.

Adding the mother's subjective well-being variables in Model 3 explains a further 5 percent of the variance in the child outcome (4 percent according to change in adjusted R^2), a significant improvement in prediction. Higher scores on the maternal Life Satisfaction measure at the time of the first follow-up predicted a more favorable perception at 42 months of the child's positive social behaviors.

While Model 4 (adding the human capital development variables) did not significantly improve our ability to predict the Positive Behavior Index, adding the larger social context variables in Model 5 explains a further 7 percent of the variance in the child outcome (or 5 percent according to change in adjusted R^2), a significant improvement. When families had moved more often, and when mothers reported that they had no social support at 42 months, their children's scores on the Positive Behavior Index were lower.

In general, although we do not see the predicted association between mother's human capital and more favorable child outcomes, findings for these two maternal report measures of the child's social behavior do support the view that maternal distress and stress in the larger social environment predict less positive social behaviors, whereas parenting behavior that is supportive and less harsh predicts more positive social behavior.

B. Findings for Teacher-Reported Behavior Problems and Positive Behaviors

1. Behavior Problems Index. As Table 8.3 shows, a substantial proportion of the variance in the teacher's report of the Behavior Problems Index is explained by the set of background characteristics ($R^2 = .19$, adjusted $R^2 = .15$). In particular, teachers reported more behavior problems for boys and for children whose mothers had lower literacy scores at baseline (see Appendix Table C.3). The final model (Model 5), encompassing all of our predictor variables, explains 38 percent of the variance in the Behavior Problems Index (adjusted $R^2 = 21$ percent), again a substantial proportion of the variance. However, as can be seen in the table, when we move from Model 1 to Model 2, from Model 2 to Model 3, and so on, no additional block of variables beyond the background characteristics can be said to result in a significant increase in our ability to explain the teacher's report of behavior problems.

2. Positive Behavior Index. Family background characteristics also play an important role in the teacher-reported Positive Behavior Index (see Table 8.4). Nineteen percent of the variance in this outcome (adjusted $R^2 = 15$ percent) can be explained by family background variables. Higher scores on the Positive Behavior Index were predicted specifically by the child's age (teachers gave higher scores to younger children), by gender (girls were given higher scores), and by higher maternal literacy scores at baseline. Those families not at the Portland site also receive higher scores (see Appendix Table C.4).

However, Table 8.4 shows that, unlike the findings for the teacher's report of the Behavior Problems Index, two further blocks of variables did add significantly to our prediction of this outcome: the parenting variables (Model 2) and the larger social context variables (Model 5). Adding the parenting variables in Model 2 explains a further 9 percent of the variance (change in adjusted R^2 is 4 percent). Interestingly, the specific variables that are predictive are both observational measures of cognitive stimulation: the Percentage of Immediate Utterances and a higher score on the Wheeled Objects Named as a Proportion of Maternal Elicitations measure. Immediate Utterances in the context of the observational study's book reading task involved naming, labeling, and discussing things directly portrayed in the book. The Objects/Elicitations measure is seen as a measure of the mother's effectiveness in providing the amount and kind of information needed to get her child to name wheeled objects. At the 42-month follow-up, with children in the observational study between about ages 4 and 7, teachers might well appreciate children's attentiveness to tasks involving labeling, naming, categorization, and responding to questions, all behaviors that could be encouraged by maternal immediate talk and effectiveness in a guessing game.

Adding the variables reflecting on the larger social context (moving from Model 4 to Model 5) also results in a significant increase in the proportion of variance in this child outcome that can be explained (a further 8 percent according to R^2, and 4 percent according to adjusted R^2). When mothers reported fewer difficult life circumstances at 42 months, teachers rated their children as having more positive social behaviors. With all variables taken into account in the final model, more than a third of the variance in teacher's report of Positive Behavior Index is explained (more than a fifth according to the adjusted R^2).

In general, much of the variance on the two teacher-reported measures of the children's social behavior can be explained by the child and family background characteristics. For the Positive Behavior Index, though not the Behavior Problems Index, we see some support for the ecological perspective: parenting but also the larger social context of the family help to predict the teacher's perception of the child's positive social behaviors.

C. Findings for the Bracken Basic Concept Scale

The pattern of findings for the direct assessment of the children's cognitive development, the Bracken Basic Concept Scale School Readiness component, is most similar to that for the teacher's report of the Behavior Problems Index, in that the set of background variables explains a significant proportion of the variance in the child outcome, but no further set of variables adds significantly to our ability to predict the outcome. As can be seen in Table 8.5, Model 1 explains a significant proportion of the variance in children's scores on the Bracken Basic Concept Scale ($R^2 = 10$ percent, adjusted $R^2 = 7$ percent). Higher scores were predicted by higher maternal literacy at baseline, having fewer children in the family at baseline, and greater child age (Appendix Table C.5). The final model, with all variables included, also explains a significant proportion of the variance ($R^2 =19$ percent, adjusted $R^2 = 7$ percent). Yet moving beyond Model 1, for none of the successive models does adding a block of variables result in a significant increase in the proportion of variance explained. Thus, it cannot be said that the blocks of parenting, mother's subjective well-being, mother's human capital, or larger social context variables significantly improve our prediction of children's scores on the Bracken Basic Concept Scale.

D. How These Findings Address Our Specific Questions

We return now to the specific questions motivating these analyses.

1. **Does more stimulating and supportive parenting predict more positive child outcomes? Does consideration of parenting variables significantly improve prediction of child outcomes?** Parenting measures added significantly to our ability to predict three of the child outcomes (both of the maternal report measures concerning the child's behavior and the teacher's report of the child's positive social behavior). These findings corroborate our more detailed analyses of parenting measures as predictors of the child outcomes in Chapter 7. More favorable child outcomes were predicted by less harsh, more warm, more cognitively stimulating, and, interestingly, more confident parenting behavior. In general, we can answer our first question affirmatively.

2. **Does greater maternal subjective distress predict less positive child outcomes? Does adding these variables improve prediction of child outcomes?** Consideration of maternal subjective well-being variables at the time of the first follow-up in the New Chance Evaluation did significantly improve prediction of two of the child outcomes. Interestingly, both of these outcomes relied on maternal report. Less favorable scores on the two maternal report child outcome measures were predicted specifically by greater stress in the parenting role and less overall life satisfaction. Thus, while we see some support for the hypothesis that maternal subjective well-being helps to shape child outcomes, the case would be stronger if the evidence extended to child outcome measures not relying (as do the subjective well-being measures) on maternal report.

3. **Does greater maternal human capital predict more positive child outcomes? Does adding these variables improve prediction of child outcomes?** There is little evidence in these analyses to indicate that consideration of variables reflecting mother's human capital (specifically whether the mother had completed high school or a GED, obtained a trade license, was currently employed, was currently receiving AFDC, and her earnings since enrolling in the New Chance Evaluation) added significantly to our ability to predict child outcomes. When the human capital variables were added to background variables, parenting variables, and mother's subjective well-being variables, in no instance did we see a significant improvement in our ability to predict to child outcomes.

This finding is inconsistent with previous research indicating that such variables as maternal educational attainment and earnings are important predictors of child outcomes. Previous research on this issue has often involved samples with greater variation on such variables as maternal education and earnings. Perhaps the much narrower range on these variables within the present sample of mothers (all with limited education and income) does not provide sufficient variability to result in significant associations with child outcomes. It may also be the case that other analyses considering the role of mother's human capital have not controlled for as many variables as we have in the present analyses (in which the human capital development variables were considered cumulatively along with control variables, parenting variables, and mother's subjective well-being variables). A further possibility is that recent human capital achievements by mothers may show positive influences on the development of children only after a period of time: our measures of mother's human capital and of children's development were derived contemporaneously from the 42-month follow-up interview. Finally, as is suggested in the 42-month report, it is possible that the variation that does occur within the New Chance sample (having obtained a GED or not, being employed or not) does not signify a major difference in the lives of the mothers and their children. Indeed, the concern is expressed in the New Chance 42-month report that the completion of the GED may mean little increase in the likelihood that a mother will find stable employment. If so, attainment of the GED may signify little change in the children's lives.

4. Does stress in the larger social context predict less positive child outcomes? Does adding these variables improve prediction of child outcomes? For three of the child outcomes examined (the mother's reports of behavior problems and positive behaviors and the teacher's report of positive behaviors), consideration of larger social context variables significantly improved prediction of the child outcome. Less favorable outcomes occurred in association with more difficult life circumstances, a greater number of residence changes, and an absence of social support, thus supporting the hypothesis that stress in the larger social context is predictive of less positive developmental outcomes. In addition, however, it is noteworthy that participation in day care was also associated with behavior problems in these children. It is possible that the children in this sample participated in day care that was not of sufficiently high quality or duration to support positive development. As we have noted in Chapter 3, even the day care provided through New Chance is not of high quality. Haskins (1985) has summarized evidence that day care participation is sometimes associated with an increase in child aggression. The possibility exists that for children whose families are facing poverty and multiple life stressors, child care of higher quality and stability, perhaps provided over a longer period of time, is needed to engender positive social development.

5. Do the same sets — or different sets — of variables serve as significant predictors of the measures of social behavior as reported on by mothers and by teachers? It is noteworthy that the child and family background characteristics explain a higher proportion of the variance in the two outcomes reported on by the teachers than in those reported on by the mothers. Beyond this, the pattern of prediction for the teacher's Positive Behavior Index is similar to that for the mother's Positive Behavior Index (except that the subjective well-being variables add to the variance explained in the mother's report but not the teacher's report version of this outcome). However, the pattern of prediction for the teacher's and mother's Behavior Problems Index is quite different. For the mother's report outcome, adding the parenting, subjective well-being, and larger social context variables all improve our ability to predict the outcomes, while for the teacher's report outcome, none of these blocks of variables improves prediction. Why might the sources or contributors be different for mother- and teacher-perceived behavior problems but not positive behaviors? One possible explanation is that children may manifest differing behavior problems at home and at school, but the two contexts may elicit similar indications of social competence. Previous research suggests that teachers may be particularly sensitive to externalizing (that is, acting out, aggressive) behavior problems, as this is disruptive to instruction and classroom routines. Parents, by contrast, may be more aware of internalizing (withdrawn, depressed) types of behavior problems, which are not as disruptive but may be of concern to mothers and fathers (Hetherington, Cox, and Cox, 1985). Patterns of prediction may differ if mothers and teachers are sensitive to different aspects of child behavior problems. Another possibility is that when the mother is the source of information, an underlying variable, such as a generally positive or negative view of her life, may be reflected not only in her report of her own subjective well-being and life circumstances, but also in her perception of her child's behavior problems.

6. Does an ecological perspective, which takes into account the larger social context in which parenting takes place, help explain our paradox — that is, that program impacts on parenting behavior do not result in positive program effects on child outcomes at 42 months? These analyses support the perspective that variables other than parenting behaviors are important to children's development. In particular, we have seen that mother's subjective well-being and the larger social context are important in addition to parenting behavior in shaping child outcomes. Those interested in enhancing children's development in welfare families with young mothers may thus need to broaden the scope of their concerns beyond parenting behavior.

While these analyses suffice to tell us that multiple factors influence children's development within our *sample as a whole* (including mothers from both the experimental and control groups), further work will be needed to address the paradox *specifically for experimental group families* that there were no positive program impacts on child outcomes despite the evidence of modest impacts on parenting behavior[4]. The present analyses do not address several key issues. We do not know, for example, whether those mothers within the experimental group who made progress in terms of their parenting behavior were the same mothers who experienced heightened depression, stress, and residential mobility. That is, we do not know whether progress in one sphere of the lives of particular mothers was overwhelmed by setbacks in other spheres. Perhaps within the observational study experimental group, rather than counterbalancing influences for particular mothers, *different* groups of mothers progressed in parenting and experienced declines in terms of such variables as depression and stress. That is, we have not learned from these analyses what the specific dynamics are for families within the experimental group that yield the paradox. We also do not know whether positive program impacts for children would emerge at some future point in time. Perhaps our findings are not so much a paradox as a matter of when children's development was assessed. Finally, we cannot generalize from the present findings to the full New Chance Evaluation sample, given that the full sample includes families from further program sites, families with children beyond the age range studied here, and Hispanic families as well black and white families.

We can use the present analyses, however, to articulate a hypothesis for the divergent parenting and child outcome findings for the experimental group families. We have seen that mothers in the New Chance Observational Study entered the study with high levels of depression and facing multiple stressors in their lives. We have seen also that mothers in the experimental group of our sample made modest progress in terms of parenting, but at the same time showed more signs of dissatisfaction with their lives and more indicators of depression. A hypothesis that could be examined in detail in future work is that child outcomes were generally not influenced, or were negatively influenced, by the New Chance Program because of counter-balancing program effects for particular mothers on parenting and on variables reflecting psychological well-being and stress in the larger social context.

---

[4] Further analyses by members of the New Chance Observational Study team addressing these issues are planned.

III. Conclusion

Taken together the findings of Part I can be seen as both encouraging and sobering. We can be heartened by the finding that the New Chance Program was able to bring about positive parenting impacts, even in a population of mothers burdened with serious life stressors. These positive program impacts were detected even though mothers generally did not enter the New Chance Program to improve their parenting behavior, and even though the dosage of parenting education was limited. The evidence of program impacts, while modest in magnitude, extends across measures of parenting that differ substantially in terms of how the information was collected and whether the focus was on the cognitive or affective aspect of the interaction. We also have evidence that participation in multiple aspects of the New Chance Program, including but going beyond parenting education, was associated with the parenting behavior of mothers in the experimental group.

Yet at the same time our results indicate that if we are to understand children's development within this sample, we need to broaden our consideration beyond parenting behavior. Supportive and stimulating parenting behavior is clearly important for the children in this sample, helping to explain child outcomes assessed nearly two years after parenting behaviors were measured. But there were other important predictors of child outcomes as well. Our findings point to the importance of maternal subjective well-being and the broader social context (including children's experience of child care) as important in shaping child outcomes.

Programs for very young mothers in poverty that seek to improve outcomes for children *as well as* for mothers need to take into account the many contributors to children's development. The possibility exists that positive program impacts on parenting, difficult to achieve and important as they are, may not translate into positive child outcomes if programs do not focus also on the families' circumstances and the quality of children's experiences in child care.

Chapter 9

Key Findings and Their Implications

Martha J. Zaslow

In Part I of this monograph we have attempted to ***describe*** *the parenting behavior of mothers in the New Chance Observational Study sample; to examine whether the New Chance Program had* ***impacts*** *on differing aspects of parenting behavior; to consider whether and how differing measures of parenting behavior* ***predict*** *child outcomes; and to ask how parenting behavior* ***combines*** *with other important factors in shaping the development of the children in this sample. In concluding Part I, we summarize key findings regarding each of these issues, place these findings in the context of previous research, and consider the implications of our results.*

In the introductory chapter of this monograph, we stated the four goals of the New Chance Observational Study. Those goals were to examine: (1) Whether a comprehensive intervention for young mothers who are high school dropouts receiving welfare improves their parenting practices; (2) How closely measures of parenting behavior collected via direct observation of mother-child interaction and through interviews with mothers are related, and how well each kind of parenting measure predicts to child outcomes; (3) What role parenting behavior plays, relative to other important influences in shaping the development of the young children of mothers participating in this intervention; and (4) What methodological issues are raised when researchers augment a program evaluation by embedding observational research in a survey research context. This chapter will attempt to summarize our findings for the first three goals (the last goal being addressed in Part II), and to discuss the implications of these results.

We lay the groundwork for a discussion of the findings on program impacts on parenting behavior (goal #1) by summarizing our results on how the various measures of parenting are related to background characteristics of the mothers and families.

I. Descriptive Findings on Parenting Behavior: Laying the Groundwork for Examination of Impacts

Earlier chapters of Part I described the parenting behavior of the adolescent mothers in the New Chance Observational Study sample. The richness of the background information available for each family in the sample provided an excellent context within which to examine characteristics associated with more and less supportive and stimulating parenting behavior within this sample.

The descriptive analyses lay the groundwork for the examination of program impacts, but should not be seen as merely preliminary to the examination of impacts and prediction of child outcomes. Indeed, the descriptive analyses we report on address significant gaps in knowledge about teenagers as mothers. The research to date describing the parenting behavior of adolescent mothers has focused almost exclusively on the period *immediately following the transition to parenthood*. Thus, we know a great deal about the interactions of teenage mothers and their newborns and infants, but little about mother-child relations beyond the infancy period. According to Chase-Lansdale, Brooks-Gunn, and Paikoff (1991), failure to go beyond the infancy period in research on the parenting behavior of teenage mothers is particularly serious in light of findings that differences in outcomes for the children of adolescent mothers begin to appear only during the preschool period (Brooks-Gunn and Furstenberg, 1986; Hayes, 1987; Moore, 1986). Furthermore, failure to go beyond the period of transition to parenthood ignores the longitudinal research on the life course of the adolescent mothers themselves. Such research indicates recovery in terms of educational and employment activities in the mothers as children grow older (Furstenberg, Brooks-Gunn, and Morgan, 1987).

By focusing on children approximately 30 to 60 months old, the New Chance Observational Study begins to fill this gap. Further, while previous descriptive studies have tended to examine variation in the behavior of adolescent parents in light of one or two background characteristics deemed to be particularly important (such as social support to the mother or residence with the family of origin), the New Chance Observational Study provides an unusual resource for descriptive work because of the wide range of maternal and family characteristics documented just prior to enrollment in the evaluation, and also 18 months into the evaluation.

How do the descriptive results presented in Chapters 4–6 of this monograph, relating characteristics of the mothers and families to parenting behavior, help to build a picture of parenting behavior among low-income adolescent mothers with preschoolers?

A. Social Support and Parenting Behavior

Previous research suggests that we attend especially to variables reflective of social support. A series of studies focusing on mother-infant interaction documents more positive parenting behavior when adolescent mothers have more such support (Colletta, 1981; Crockenberg, 1987a, 1997b; Unger and Wandersman, 1988). Although there is consensus that social support is related to the quality of parenting by very young mothers, there is some disagreement as to the particular source or aspect of support that is most important (for example, support from partner or from family members; total support available or subjective sense of emotional support).

Our descriptive results point to associations of mothers' parenting behavior and their *concurrent* social support.[1] The fact that there was little indication of an association between parenting behavior and baseline measures of social support underscores the importance of the current social context. Mothers who reported more sources of social support at the time of the 18-month follow-up had more positive relations with their children not only in terms of the observed affective quality of mother-child interaction (with higher Quality of Relationship, Child's Affection to Mother, and Child's Compliance ratings), but also in terms of the literacy-related aspects of interaction (with higher overall Book Reading Quality ratings). Mothers with more sources of support also used less harsh disciplinary practices as documented on the HOME-SF Harsh Discipline subscale.

---

[1] Indeed, baseline measures of maternal satisfaction with social support were negatively correlated with two observational measures of the affective quality of mother-child interaction, though positively with the Maternal Warmth measure.

B. Maternal Psychological Well-Being

In Chapter 3 we noted that field representatives and parenting class educators, when interviewed about the parenting education component of the program, pointed to depression and low self-esteem as obstacles to positive parenting behavior among the participants in the New Chance Program. Further, we noted that stress and depression were serious problems in this sample. For example, 51.4 percent of the mothers in the New Chance Observational Study sample were at some risk of depression upon enrolling in the evaluation, a far higher proportion than is found in the general population.

We see ample evidence in our descriptive results of associations between parenting behavior and markers of the mothers' psychological well-being from the 18-month follow-up. Being at high risk for depression at 18 months was associated with lower HOME-SF scores across all subscales except that involving the physical environment. Further, mothers at high risk of depression had lower Book Reading Quality scores, reported more parenting stress, and described themselves as using more controlling disciplinary practices.

Other aspects of maternal psychological well-being were also related to parenting behaviors. For example, mothers with a greater sense of efficacy with regard to events in their lives (mastery) were observed to engage their children in richer book reading discussions and to rate higher on Ease of Ideas on the wheels task, while their children were observed to be more successful at guessing objects with wheels. Greater sense of mastery was associated with less parenting stress and less controlling discipline.

It is interesting to note that while mastery was related to varying markers of parenting behavior, our measures of mother's age and of mother's age at the time of her first birth were less consistently associated with parenting behavior, linked significantly only with a few interview-based measures of parenting. Quint and Egeland (1995) emphasize the importance of considering teenage mothers' own progress through adolescence and toward greater autonomy as co-occurring with, and perhaps hindered by, the early transition to parenthood. They note that variables indicative of ego development and maturity may be more important than absolute age or age at transition to parenthood in predicting the quality of mother-child relations among adolescents. If we see mastery as a possible marker of the young mothers' emerging sense of autonomy, then our descriptive results provide some support for the view that developmental age more than absolute age or age at first birth may be central to observed and reported parenting behavior among adolescent mothers.

C. Residing with Parents and Partners

Another key variable highlighted by previous research is that of residing with the maternal grandmother. Recent research by Chase-Lansdale, Brooks-Gunn, and Zamsky (1994) involving observations of young African American mothers and their 3-year-olds, and grandmothers of the same children, documents better quality parenting as well as grandparenting when mothers and grandmothers were *not* co-residing. It is noteworthy that this study, unlike much of the previous research, focused on children in the preschool period. We also note that the mothers in the sample, while all young, were not all adolescents at the birth of their first child.

Our descriptive findings provide little indication of differences in parenting behavior according to whether or not the mother was living with one of her parents at the time of the 18-month follow-up. No significant differences beyond what might be predicted by chance were detected for observational or interview-based measures of parenting between those mothers who were or were not residing with a parent.

We did, however, see some differences in parenting behavior according to whether or not mothers were living with a boyfriend or a husband. Interestingly, whereas the correlates of residence with a partner or husband *at baseline* were generally negative (such residence was associated with lower HOME-SF scores on several subscales), they were mostly positive when residence *at 18 months* was considered. Women who were living with a husband or a partner at the time of the 18-month follow-up had higher scores on the HOME-SF Cognitive Stimulation subscale and reported less parenting stress and less use of controlling disciplinary practices, though they also rated themselves less favorably in terms of the expression of warmth to their children. The contrasting correlates of the baseline, as opposed to 18-month follow-up, measures of residence with a boyfriend or a husband raise the possibility that such cohabitation may have differing implications for parenting according to the age of the mother (or child) when it occurs.

In both the full New Chance Evaluation sample and the observational study sample, group differences were found in the residence patterns of the young mothers at the time of the 18-month follow-up: Mothers in the experimental group were less likely to be living with a parent or grandparent. The pattern of program impacts on residence taken together with our descriptive results on parenting behavior and the findings of Chase-Lansdale and colleagues suggest the need for further work exploring residence patterns among adolescent mothers. We need to learn more about the implications of residence patterns for parenting behavior, as well as about the context in which these patterns come about. For example, would further research reveal differences on measures of maternal psychological well-being for young mothers residing with parents, with partners, or independently? Would further research reveal indications of conflict between young mothers and their own mothers when they are co-residing? Should independent residence or residence with a partner be seen as part of progress toward autonomy, or does it foreclose such progress? [2]

D. Education and Literacy

In accord with previous findings that maternal education and literacy are strong predictors of mother-child interaction and of child outcomes (for example, D'Amico, Haurin, and Mott, 1983; Desai, Chase-Lansdale, and Michael, 1989; Moore and Snyder, 1991), our baseline measures of maternal educational attainment and literacy were associated with more positive parenting behavior, particularly as manifested in the HOME-SF total score and subscales, and through observational measures of literacy interactions (for example, Book Reading Quality ratings). By contrast, the baseline measures of previous employment and of whether the mother had grown up in a family that received AFDC were not predictive of our parenting measures.

---

[2] Further work examining the correlates and implications of differing residence patterns within the New Chance Observational Study sample is planned by Deborah Coates.

E. Participation in Child Care

We note with interest that observed mother-child literacy behaviors as well as the affective quality of interactions were more positive when the child had participated in any child care during the 18-month follow-up period. Children who had participated in child care named more objects during the wheels task and were more responsive to their mothers' hints during this task. Mothers whose children had participated in some child care had higher ratings on Book Reading Quality and engaged their children in more complex discussions following the book reading. Children who had participated in child care were observed to be more compliant, and their mothers showed less harsh discipline.

The fact that both maternal and child behavior differed when children had attended some child care raises the possibility of child effects: children who have attended some child care may show greater interest in book reading and word games, as well as greater compliance with adult requests, because of expectations and activities they have experienced in child care. The children's greater interest and compliance may, in turn, elicit more complex and stimulating maternal behavior and require less harsh disciplinary tactics. Another possible interpretation focuses on self-selection: there may be differences between families who do and do not choose to use child care. For example, mothers who choose to use child care may place a higher priority on the provision of stimulation to their children, a priority that may be reflected in patterns of mother-child interaction in the home as well as participation in child care.

It is important to note that while child care participation up until the 18-month follow-up was associated with more positive parenting behavior, our subsequent analyses (reported in Chapter 8) document that child care participation through 18 months ultimately predicted less favorable development on some child outcomes. The children in the experimental group within our sample showed a pattern of entering but then leaving day care in association with their mothers' participation (and then completion) of phase 1 of the New Chance Program. The possibility exists that those children who participated more in day care settings also experienced more instability in terms of their daily care situations. The more concurrent associations of day care participation and parenting behavior may reflect positive concurrent or recent day care experiences, whereas the longitudinal associations with development may be rooted in a more frequent experience of child care instability.

We also note that a special study of the quality of child care provided at selected New Chance sites, summarized in the 18-month report, found that these child care settings received an average rating of just below "good" on the Early Childhood and Infant Toddler Environment Rating Scales (Harms and Clifford, 1980). Further, on-site child care was not available at all sites, and the quality of the off-site child care that mothers used for their children was not documented. Previous research suggests that positive developmental outcomes for children in high-risk samples are associated with sustained and substantial participation in high-quality child care. Thus, the child care available through the New Chance Program may not have been of sufficient quality or duration to support positive developmental outcomes.

In sum, our descriptive findings are in accord with previous results pointing to the importance of social support, maternal psychological well-being, education, and literacy as predictors of parenting behavior among adolescent mothers. Although our findings do not indicate less positive parenting behavior among young mothers who co-reside with a parent, they do underscore the importance of continuing to examine the role of residence patterns in the lives of young mothers. Finally, there are interesting indications in our descriptive findings that children's participation in child care was associated with more positive maternal as well as child behavior during interactions. We need insight into why child care participation then predicted less favorable child outcomes on some measures at the time of the 42-month follow-up.

II. Impacts of the New Chance Program on Parenting Behavior

Most intervention and evaluation studies for adolescent mothers and their children have aimed to enhance either the economic circumstances of the mothers or the school readiness of the children (Chase-Lansdale, Brooks-Gunn, and Paikoff, 1991). The New Chance Demonstration was one of a small group of two-generation programs: that is, interventions that seek to improve the circumstances of both mother and child simultaneously (Smith, 1995).

The New Chance Observational Study sought to provide a detailed examination of program impacts on various aspects of parenting behavior, complementing interview-based measures of parenting available for the full evaluation sample with sensitive observational measures and several further interview measures for a subset of the larger evaluation sample. We know of only a handful of studies involving a detailed examination of impacts on parenting behavior within two-generation interventions. Accordingly, the New Chance Observational Study is helping to build a new literature. What has been learned about program impacts on parenting *within* the New Chance Observational Study? How can these findings be related to findings from other studies reporting on parenting behavior in two-generation interventions?

A. A Summary of Impacts on Parenting Behavior Within the Observational Study

We have noted from the outset that there are both reasons to expect the New Chance Program to have positive effects on parenting and reasons to exercise caution in our predictions. On the one hand, New Chance sought directly to affect the parenting behavior of the young mothers in the program through a parenting education component. Guidelines for parenting education within New Chance were developed with the guidance of an advisory board of developmental psychologists with special expertise in the area of teenage childbearing and parenting. In addition, the parenting education program was tailored to the needs of young low-income women.

The New Chance Program also touched on parenting issues in program components other than parenting education. For example, life skills classes provided guidance on assertive-but-positive communication skills within the family, and family planning classes included discussion of the importance of birth spacing to the time and attention available for each child. Further, the quality of parenting could have been improved through the stimulation that mothers received in preparing for employment roles or through overall contact with this comprehensive and supportive program.

Yet we have also noted that the mothers participating in New Chance were heavily burdened by major life stressors, such as housing instability, history of emotional or physical abuse, and dysfunctional parenting in their own families. The young mothers volunteered for New Chance not primarily for the parenting education component, but rather for basic education, and especially training for the GED examination. Further, participants in New Chance did not always feel that they needed parenting education. When they did attempt to change their parenting practices, the mothers sometimes experienced resistance or undermining behaviors from partners and family members. Field representatives and parenting education instructors felt that many of the young mothers faced serious psychological difficulties, such as depression and lack of self-esteem, that impeded their efforts to be better parents. They also felt that the principles and positive practices that they sought to convey in parenting education were harder for this group of mothers to absorb than they had expected.

The dosage of parenting education was also somewhat limited within the New Chance Program. First, approximately one-third of mothers in the experimental group, by their own report, never attended New Chance parenting classes, and nearly one-fifth of mothers in the control group participated in community-based parenting classes. Further, mothers in the experimental group within the observational study sample spent an average of about 18 hours in parenting education. By contrast, the average number of hours spent in basic education was about 100.

Looking across the findings on program impacts for all of the parenting measures included in the New Chance Observational Study, we find that our summary statement must closely parallel our statement of expectations above. That is, our findings must be summarized in a format of "on the one hand" and "on the other hand."

On the one hand, as we noted in Chapter 7, for each type of parenting measure included in this study (observational measures of literacy interactions, observational measures of affective quality, HOME-SF subscales, Maternal Report scales, and Time Use measures), we have documented at least one significant group difference indicating more positive parenting behavior in the New Chance experimental group. Mothers in the experimental group had higher scores on observed Book Reading Quality and lower scores on observed Harsh Treatment of the focal child. HOME-SF subscale scores indicated that mothers in the experimental group provided a higher quality home environment overall, with a significant difference specifically in the area of emotional support provided to the focal child. Mothers in the experimental group perceived themselves to be warmer in their interactions with their children and to spend more time with their children, particularly in carrying out parenting chores.

There was a single significant exception to this pattern of positive impacts: Mothers in the experimental group were rated lower on Ease of Ideas when asked to carry out a verbal guessing game with their children. However, this difference was found to be attributable to a larger subset of experimental group than control group mothers who did not grasp the goal of this task. Thus, this group difference might just as readily be seen as an indication that task or test anxiety was evoked by this mother-child task (which was the only one not supported with physical props and could perhaps be seen by the mothers as test-like) or as an indication of problems in interviewer performance in explaining or structuring the task.

Positive program impacts have been documented in the New Chance Observational Study for measures relying on differing informants (mother, interviewer, coder), across measures differing in breadth (seeking to document only dyadic interactions of mother and child or the input of further social partners and the home environment), and across measures focusing on cognitive as well as socioemotional aspects of mother-child interaction. This consistency of impacts across types of parenting measures increases our confidence that New Chance affected parenting behavior positively.

On the other hand, our findings cannot be seen as pervasive, given the number of discrete measures examined. For example, only one of eight measures of mother-child literacy interactions was found to differ significantly by group. When we examined effect sizes for the observed program impacts, all but one fell in the range indicating small program impacts. Only the difference on the HOME-SF Emotional Support subscale was of moderate size.[3] Further, only this program impact on the HOME-SF was significant across all major population subgroups. In sum, although we have documented positive program impacts across a range of different parenting measures, these impacts are modest in terms of magnitude, number of variables for which impacts were detected, and consistency of significant effects across families with differing initial characteristics.

B. The Contribution of Program Impact Findings to the Literature on Two-Generation Interventions

How do our findings contribute to the emerging literature on parenting behavior in two-generation interventions? Two further two-generation interventions have recently reported findings for observational studies of mother-child interaction. An observational study was embedded within the evaluation of the Teenage Parent Demonstration (Aber, Brooks-Gunn, and Maynard, 1995), an intervention for teenage mothers on welfare that barely qualifies as a two-generation intervention in that it included only minor program components addressing the needs of children. The evaluation of Avancé, a more fully two-generational intervention for low-income Hispanic families, also included a detailed evaluation of program impacts on parenting (Walker et al., 1995).

Although our findings can (and should) be related to those from these two other studies, we must exercise caution in comparing findings across studies. As Smith (1993, 1995b) has noted, the two-generation interventions that are now being studied differ on several dimensions simultaneously. For example, they target differing populations for participation and may enroll families as volunteers or on a mandatory basis. The programs themselves differ in terms of comprehensiveness, intensity, and duration.

It is tempting to assume that differences in findings on parenting behavior across studies are attributable to the presence and dosage of particular program components, especially parenting education. Yet differences in findings across studies may in fact be related to sample differences or to differences in overall program comprehensiveness or supportiveness. Smith (1993) proposes that as the literature on two-generation interventions matures, researchers move toward implementation of planned variation studies to address this problem.

---

[3] As discussed in Chapter 7, our examination of effect sizes focused on five of the seven parenting measures for which positive program impacts were found (Book Reading Quality, Harsh Treatment, HOME-SF Emotional Support, Maternal Warmth, and Parenting Chore Time). In the two instances where both a summary measure and a more specific measure were found to show positive program impacts (HOME-SF total and HOME-SF Emotional Support measure; Overall Parenting Time and Parenting Chore Time), we examined the effect size only for the more specific measure.

Keeping this caveat in mind, we note marked differences in findings of impacts on parenting behavior across three studies to date of two-generation interventions that involve evaluations of program impacts on parenting behavior. The observational study embedded within the Teenage Parent Demonstration found *no program impacts* on mother-child interaction, videotaped, as for the present study, in the home during a mother-child task. For New Chance we have noted *modest positive program impacts scattered across a number of different parenting measures.* Findings for Avancé point *to consistent positive program impacts on parenting behavior.* In addition to a difference on the HOME total score two years after program enrollment, mothers participating in Avancé were observed to show more positive affect, social initiation, vocalization, contingent praise, and encouragement of the child's verbalization one and two years after enrollment. They also spent more time in teaching behaviors and showed a higher quality of communicative verbalizations. Dyadic interaction between mothers and children for those in the program group showed more mutual enjoyment, joint attention and activity, and reciprocity.

In order to place these findings in context we need to understand not only how parenting education was approached in each intervention, but also how samples were selected, what other program components were involved, and how comprehensive and intensive each program was. In the Teenage Parent Demonstration, all adolescent mothers entering the welfare system in specific sites (Camden and Newark, New Jersey, and Chicago) were assigned randomly to receive either regular or enhanced services. Those receiving enhanced services could receive their full welfare grants only if they actively pursued programs to enhance their education or employment skills. Financial sanctions were applied for nonparticipation. Mothers in the enhanced services group also received such supportive services as child care assistance, parenting and life skills workshops, opportunities to link education with job training and employment experiences, and case management. The parenting intervention component of this program involved an initial workshop session followed by case manager assistance to individual mothers on parenting issues as needed.

Avancé was a two-generation program serving low-income Hispanic families with young children. Participation was voluntary, and a major reason that mothers entered this program was their interest in parenting education. A first program phase involved attendance at a three-hour parenting class once a week over the course of nine months. Those who opted to participate in the second phase of the program focused on their education and employment skills. Although there were other program components within Avancé (for example, advocacy for the families in locating assistance and other service programs, home visits, child care while mothers attended the program), this was a far less comprehensive program than New Chance.

It is indeed tempting to view the progression across these three studies from no parenting impacts (Teenage Parent Demonstration) to modest parenting impacts on specific measures (New Chance) to consistent positive impacts across measures (Avancé) as simply reflective of the increase in the dosage of the parenting education component apparent across the three interventions. Yet, it is apparent that other key program features differentiate these programs. Above all, the Teenage Parent Demonstration was mandatory, enrolling all teenage mothers applying for welfare. By contrast, New Chance and Avancé enrolled mothers who themselves came forward and volunteered; these mothers may have been at a point when they were generally motivated to progress in their lives. Perhaps positive program impacts on parenting occur when mothers show general readiness to address problems in their lives and receive positive program inputs of various kinds to do so. Interestingly, in Avancé, a program for which mothers volunteered primarily for the parenting education component, parenting impacts were strongest, which could reflect this specific motivation to improve parenting behavior among volunteering mothers.

New Chance and Avancé , the two programs for which some positive program impacts on parenting were documented, were highly supportive in their intervention approaches in addition to being voluntary. While case workers in the Teenage Parent Demonstration sought to be supportive, this was not the only approach used with the young mothers. Mothers in this program could be sanctioned for nonparticipation.

In sum, we see several hypotheses emerging from this set of observational studies within two-generation interventions. Future research will need to explore the possibilities that parenting impacts are more likely to occur in such interventions when (1) programs have sufficient dosages of parenting education or other central program components that can affect parenting behavior; (2) programs enroll mothers who show some initial motivation to improve their life circumstances, or more specifically their parenting behavior; and/or (3) programs provide mothers with social support.

As the research on two-generation interventions progresses, a further issue should be addressed: the possibility of indirect effects on children of program components directed not at the children or aimed at enhancing parenting, but rather that seek to enhance family economic self-sufficiency (Chase-Lansdale, Brooks-Gunn, and Paikoff, 1991). In keeping with this possibility, our examination of program components associated with parenting behavior in the New Chance experimental group pointed to the importance not only of program components with content specifically related to parenting behavior, but also of the human capital development components of the program and of participation in the program overall. We are left with the intriguing question of why greater participation in program elements like employability training and work internships might enhance parenting behavior in a voluntary program like New Chance (though not in the Teenage Parent Demonstration).

Our analyses linking program participation and parenting behavior controlled for those baseline characteristics associated with extent of program participation. That is, we controlled for the *observable* or *measurable* aspects of self-selection for greater program participation. Yet it is possible that further maternal characteristics, not measured at baseline, were associated with program participation. For example, we had no baseline measures of motivation to succeed in a program like New Chance. Perhaps such an unmeasured characteristic underlies both the greater program participation (and especially participation that continued into the human capital development components of phase 2 of the program) and more supportive and stimulating parenting behavior.

A second possible interpretation of the link between participation in human capital development program components and parenting behavior comes from the research on maternal and paternal employment. Researchers have hypothesized that specific features of mothers' and fathers' jobs come to influence their childrearing beliefs and practices. Building on the work of Kohn and colleagues (for example, Kohn and Schooler, 1982), Menaghan and Parcel (1995), for example, propose that jobs that are repetitive, unstimulating, and offer little opportunity for self-direction may be associated with childrearing values that emphasize obedience to adults. By contrast, when jobs involve greater variety, stimulation, and self-direction, parents are more likely to use strategies of reasoning in disciplining their children and to expect self-direction from their children in their behavior through internalization of adult norms. Menaghan and Parcel also speculate that more cognitive stimulation of the parent on the job may result in more varied and interesting parent-child interactions.

Perhaps a parallel process occurs in a program such as New Chance as mothers prepare for employment. That is, participation in employability development and job skills training may suffice to communicate to, and encourage in mothers, those behaviors that will be valued in employment situations. New Chance sought explicitly to prepare mothers for jobs with some potential of advancement, thus involving some degree of complexity and stimulation. Such behaviors as restraint in interpersonal interactions, work effort, and punctuality on the job may all be described and encouraged in employability and job skills training classes in programs that work to prepare low-income mothers for better jobs. As important behaviors for the mothers' employability are communicated, parallel changes may occur in terms of which behaviors mothers encourage and discourage in their children. We note, however, that the Teenage Parent Demonstration, with its strong focus on human capital development activities, had no parallel implications for parenting behavior. Perhaps such a pattern occurs only among those mothers already motivated to make changes in their lives.

Researchers may explore this possibility further through studies seeking to document which specific behaviors are encouraged in mothers during courses focusing on human capital development within two-generation interventions, and through analyses exploring the linkages between such course content and parenting behavior.

III. Differing Measures of Parenting Behaviors as Predictors of Child Outcomes

In the concluding chapters of Part I, we asked how well different types of parenting measures, assessed about 18 months and 21 months after mothers enrolled in the evaluation, functioned as predictors of child outcomes measured about 42 months after enrollment. Our examination of different kinds of parenting measures indicated that parenting measures that relied partly or completely on the perspective of an outside observer (the HOME-SF subscales, which rest on both maternal report and interviewer ratings, and the observational measures) were stronger predictors of the child outcomes than parenting measures that relied entirely on maternal report. In a stringent examination of the issue of whether observational variables add to our ability to predict child outcomes even with all the other parenting measures available to us already taken into account, we found that this was indeed the case for two (of the five) child outcomes we considered. Thus, we have evidence that observational measures of parenting can enhance our ability to explain variation in children's scores for some child outcomes.

We have been careful to note the caveat that our results reflect on the *specific* parenting measures included in the present evaluation. Our conclusions might be different if, for example, a more exhaustive set of maternal report measures had been used. If the pattern of results found here is replicated in other studies, there would be important implications for the selection of parenting measures in future program evaluations. Despite the additional cost and difficulty of carrying out in-home interviews (that permit the collection of interviewer rating data) and direct observations of mother-child interaction, these may be important resources for evaluation studies with a focus on parenting behavior and children's developmental outcomes.

IV. The Role of Parenting and Other Important Factors in Shaping Child Outcomes

When we broadened our examination of predictors of child outcomes to variables beyond parenting measures, an important pattern emerged. Warmer and more stimulating parenting behavior did predict more positive developmental outcomes within our sample. Yet other variables also served as significant predictors of child outcomes. In particular, we found that variables reflecting maternal psychological distress (for example, greater parenting stress, less life satisfaction) and stress in the larger social context (for example, a greater number of difficult life circumstances, more changes of residence) predicted less favorable developmental outcomes. Thus, parenting behavior must be seen as *combining* with other key factors in shaping children's development.

These findings are very much in accord with the urging of researchers taking an "ecological perspective" (for example, Bronfenbrenner, 1979, 1986). If we are to understand the factors shaping children's development, we must look not only *within* the family, but also at the broader *context* of the family. Features of the broader social context may be of particular importance to consider in families facing multiple serious life stressors.

Our findings carry two critical implications for those seeking to enhance the development of children of young welfare mothers. First, it is possible to improve the quality of parenting behavior, even in a sample of young welfare mothers facing major obstacles to improving their lives. Second, while mother-child interaction is an important factor in shaping child outcomes, it does not function alone. Parenting behavior combines with other key influences, especially mothers' psychological well-being, families' living circumstances, and children's experiences in child care. In order to enhance the development of the children of young welfare mothers, it may be necessary to target not only the quality of mother-child interaction, but also these aspects of the families' broader social context.

Part II: Methodological Assessment of the New Chance Observational Study

Chapter 10

Expanding the Methodological Horizons of Child Development and Survey Research

Carolyn A. Eldred

This chapter conceptualizes the New Chance Observational Study as part of an ongoing evolution that is incorporating nontraditional measures and measurement processes into surveys and making such survey capabilities accessible to a wider audience, including developmentalists. The study affords a unique opportunity for examining methodological issues in the measurement of parenting and child outcomes, both because of the challenge involved in exploiting the survey model by having lay survey interviewers administer an observational protocol and make their own substantive ratings and because of the diversity of measures and data sources included in the study. This chapter presents the goals of Part II: to make readers aware of the strengths and limitations of the "survey model," to document how the Observational Study was conducted, to assess the success of the effort, and to consider both specific recommendations for future work and broader implications for research design.

The remainder of this chapter discusses the survey model as an option for research on parenting and child outcomes. It first defines survey research as a method involving a strict division of labor between those who collect the data — lay survey interviewers — and the researchers they represent. It then describes the survey model from three key perspectives: its production orientation, the contrived and stylized format of the survey interaction, and the programming of interviewer behaviors and standardization of procedures. Although these aspects of the model may place constraints on what is possible through survey research, the chapter goes on to highlight the contribution that 50 years of methodological research by the survey community can make to a study like this in terms of design and evaluation of measures, management of interviewer effects, and conceptualization of challenging survey efforts. The discussion concludes with a recommendation that, as developmental psychology looks to the survey community to leverage larger and sounder samples, developmentalists both recognize the constraints of the survey method and develop measures and measurement processes that build on the survey field's accumulated body of knowledge.

I. The Rationale for Examining the Study from a Methodological Perspective

The New Chance Observational Study lies at the confluence of rising interest in policy-relevant research among developmental psychologists, interventions focusing on two generations in a family, and increasing demand for nontraditional forms of survey research. The burgeoning interest in policy-oriented research among developmentalists (see Bullock, 1995) has prompted concern with the size and representativeness of the samples traditionally used in developmental psychology. One expression of such an interest in larger and more representative samples has been the increasing use of contracted survey research to collect data on parenting and child development. This has taken the form of querying parents about their parenting practices and children's development (the most traditional of the methods), having lay interviewers make ratings of parent-child interactions and the home environment, and using interviewers to administer structured parent-child interactions similar to those usually carried out in the child development laboratory.

While New Chance is by no means unique in its use of survey methods to study child development, it is one of a few studies to use survey interviewers to conduct observational work with mothers and children.[1] As an intervention with an explicitly two-generational focus, the New Chance Evaluation was designed to include a variety of survey-based measures of parenting and child outcomes obtained through maternal self-report, interviewer ratings, child assessments, and reports of teachers. To enhance the picture provided by these data sources, observational data, based on the interaction of mother and child during a structured set of activities, would be a powerful adjunct. Would it be feasible to embed an observational study within the larger evaluation and supporting survey effort?

Child development's increasing reliance on the survey method — and the New Chance study itself — has occurred at a time when the survey mission in general has been expanding to take on more varied and demanding roles. Researchers who design data collection instruments and procedures, survey interviewers who gather the data, and the survey research organizations that oversee the work are all being asked to do more. Research efforts asking more of the survey method have been imbued with an optimistic "can-do" attitude, perhaps leading all involved to underestimate the significance of each additional level of responsibility that survey organizations and interviewers have been asked to assume. A case in point was the translation of the Home Observation for Measurement of the Environment (HOME) Inventory for administration by survey interviewers, initially for the National Longitudinal Survey of Youth-Child Supplement (NLSY-CS) (Baker and Mott, 1989). The short form of the HOME developed for survey administration required survey interviewers not only to obtain self-report data from mothers but also to make their own substantive ratings of mother-child interaction and the home environment, a responsibility that represented a major step beyond simply asking questions of adult respondents and recording their answers. Having survey interviewers administer cognitive assessments to children (as was done for the NLSY-CS and the evaluations of the New Chance and JOBS Programs) expanded interviewers' usual assignment in still other ways, by requiring both that they administer psychological tests correctly and that they work effectively with children. This experience in having survey interviewers collect parenting and child outcomes data may have led naturally to the decision of several teams of researchers, at about the same time, to turn to the survey method for administration of an observational protocol. With this final expansion, researchers may need to pause to take stock of how much the survey method can realistically be expected to deliver; such an assessment might also ask how successfully the previous expansions of the survey role (for example, interviewers without training in child development making substantive ratings of mother-child interactions) have been incorporated into survey practice.

---

[1] Such work has also been done by Mathematica Policy Research as part of its Teenage Parent Demonstration (Aber, Brooks-Gunn, and Maynard, 1995), in a study of preadolescents in which discussions were videotaped (Reiss et al., 1994), and by the authors of this volume for the Child Outcomes Study being conducted by Child Trends as part of MDRC's evaluation of the Job Opportunities and Basic Skills (JOBS) Program. We are also aware of additional efforts in the planning stages.

Part II of this monograph is intended to contribute to the ongoing evolution that is incorporating nontraditional measures and measurement processes into surveys and making such survey capabilities accessible to a wider audience, including developmentalists. It seeks to contribute to these goals in three ways.

First, this chapter introduces survey research[2] and its methods to social scientists who have not made survey research the primary focus of their research careers. We offer this introduction not only to provide context for the substantive discussion of measures in Part I and the methodological discussions in Part II, but also to begin to bridge the communication gap between survey researchers and their colleagues — even from the same disciplines, such as psychology or sociology — who may emphasize a substantive research agenda more than a method. Those who have not previously contracted for survey work may find themselves unprepared for the implications of a model that involves a strict division of labor between researcher and data collection personnel and that depends heavily on lay interviewers, standardization, and a production orientation. Without a full appreciation of such constraints, researchers who commission survey work may have difficulty managing it and even find themselves disappointed with the results. At the same time, these researchers — and even others who are regular "consumers" of survey data — may be largely unaware of the field's emphasis on understanding and honing the survey measurement process itself. Survey research is, after all, a method, and so it should not be surprising that its practitioners have spawned a comprehensive methodological literature. Survey research's preoccupation with threats to the integrity of a survey (that is, things that can go wrong) is illustrated by the field's preferred language, which emphasizes "error."[3] With respect to survey measurement, survey researchers see a wide gulf between establishing the objective of a measure and designing the measure itself (Fowler, 1995), finding the latter process to be far from straightforward and deserving of extensive attention. Thus, those who commission survey work will find themselves shortchanged if they look to the field primarily for a data collection capability and fail to heed its lessons on survey-based measurement.

A second purpose of Part II is to build on the discussion in Chapter 2 of Part I to document how the New Chance Observational Study was designed and implemented. This documentation, which appears in Chapter 11, is intended to help both developmentalists and survey researchers to conceptualize and implement future work.

---

[2] Throughout this monograph, the discussion of survey research assumes performance of the work by an organization that derives a substantial part of its business from social science and/or policy-related research, generally for sponsors in academia, government, and the not-for-profit sector. This type of survey research, rather than political polling or marketing research, for example, is the focus of Part II.

[3] Survey research tends to use the language of statistics as a framework for assessing quality, focusing on "variance," "error," and "bias," although a major competitor within the field is the language of psychometric theory, focusing on reliability and validity (Groves, 1987). Survey research is concerned with two "families" of error: (1) error associated with nonobservation, such as noncoverage, sampling error, and nonresponse, and (2) error associated with observation or measurement, such as that related to the questionnaire, interviewer, respondent, mode of data collection, or interaction among these.

A third purpose of Part II is to assess the design and implementation of this challenging effort and to draw lessons for future work. This assessment is carried out in the spirit expressed by Bradburn (1985), who said: "As more is being asked of surveys and as they are more and more widely used, it is inevitable that greater attention will be paid to the quality of surveys and sources of variability in survey measurement."

II. The Survey Model as an Option for Research on Parenting and Child Outcomes

In developmental psychology laboratories, research protocols are typically administered under the supervision of a principal investigator, usually by graduate students conversant with the principles underlying the research. When data are collected through home visits, these visits are usually an extension of the laboratory, with data collection carried out by the principal investigator or those working closely with him or her. (The original form of the HOME Inventory, for example, was administered in this way.) In both types of situations, the data collectors' backgrounds and familiarity with the research constitute resources that can be tapped to help with any judgments that may be required in the course of data collection. In addition, if those collecting the data are expected to make substantive ratings, laboratory-based work would generally incorporate procedures to establish satisfactory interrater reliabilities.

But some developmentalists are departing from this familiar model to place their data collection in the hands of the survey community. What are these researchers buying when they contract for survey data collection? And perhaps more important, what do they think they are buying? Of course, researchers are seeking a data collection capability, particularly the ability to work with large, dispersed samples; such a capability would include access to a field staff and an administrative structure necessary to support a successful data collection effort. But are researchers who turn to the survey community aware of some distinctive features of the survey model that may impose constraints on their work? And are they aware of the considerable methodological resources offered by the field in addition to a data collection capability?

A. What Is Meant by the "Survey Model"

Survey research employs a method in which the work of "absent researchers" is placed in the hands of distant proxies, that is, nonresearchers who conduct survey interviews. These survey interviewers are rewarded for production (with an emphasis on high response rates, cost-effectiveness, and schedule), in addition to accuracy in data collection. In order to perform these central aspects of their work, they must typically develop certain specialized skills, such as the ability to establish rapport quickly, obtain respondent cooperation, and even locate sample members who are hard to find, in addition to being able to administer (often complex) questionnaires. Individually, survey interviewers vary in background, in their degree of commitment to survey interviewing as a way of life (for many it is part-time or secondary to other work), and, like those in other fields, in their competence at particular aspects of their jobs. As a group, however, survey interviewers are neither assumed to possess nor expected to master a conceptual appreciation of the research in which they participate. This is simply not their role within the survey model, regardless of their individual backgrounds. In addition, survey interviewers' work may involve repeated turnover of assignments and study topics, effectively discouraging thought about the research agenda underlying any particular study, about which they have limited information in any case. (Viewed from the perspective of concern about experimenter effects, this distance from the thinking behind the research is a strength of the survey model.) For these reasons, the survey convention is to provide interviewers with explicit rules to direct them through the data collection — rules that, in theory at least, require minimal judgment to apply. As discussed below, these realities of the survey method yield a model characterized by an orientation toward production, a contrived and stylized format for the interaction of interviewer and respondent, and precise programming and standardization of interviewer behaviors.

1. **A Production Orientation.** Whether for profit or not, academically based or free-standing, survey research contractors function as businesses. They are characterized by a production orientation — a "get-the-job-done" tradition, including the need to produce high response rates on schedule and within budget. Indeed, a cost-effective perspective coupled with an ability to get the job done are reasons that researchers turn to survey contractors in the first place for help with large studies. However, the focus on production can lead to compromises that might not take place in the laboratory. These compromises occur not only because the survey environment differs from that of the laboratory, but also because different research traditions can co-exist *within* survey organizations (see Dillman, 1995). Thus, the culture and values of operations staff responsible for "bringing in a study" may clash with those of the survey researcher directing the work — whose thinking may parallel that of the researcher sponsoring the survey but who may have limited influence over day-to-day operational decisions.

2. **The Survey Interaction: A Contrived and Stylized Format.** Those who are tempted to think that designing questionnaires and data collection procedures should be simple and straightforward tasks for anyone educated in the social or behavioral sciences would do well to consider the fact that the survey interaction is contrived and highly stylized. These are features of the survey interaction that survey researchers have long recognized but that those outside the field may not have thought much about. Although the survey interaction has many of the hallmarks of a conversation, it is clearly not a conversation. First, and foremost, the author of the questions that are being asked is not present. Second, because of the absence of the author, the series of questions that characterize the survey interview must be entirely preprogrammed; while the questions the interviewer asks may vary to some extent depending on the respondent's answers, these paths through the interview are also preprogrammed[4] through "skip patterns." Last, other conventions of everyday conversation, especially the give-and-take that clarifies meaning and ensures understanding, are typically eliminated. As noted by Clark and Schober (1992), in everyday conversation the speaker is speaking on his or her own behalf, extemporaneously, and in interaction with the other party — using what is said to decide what to say next, within the context of whatever common understanding they share. The survey interview differs from day-to-day conversation in all of these respects.

Consider the interaction in the box below, recorded on videotape during a pretest interview for the New Chance 18-month follow-up survey. While the dialogue may have an absurd quality to those outside the field of survey research, it offers two lessons to those interested in embracing survey methods. First, the interviewer did precisely what is required by sound survey practice; she had not been provided a definition of "push or pull toy" and was not permitted to create her own. Second, this example illustrates that a measure is more than an item in a questionnaire (how many push or pull toys does the child have?). A measure consists of the entire measurement process, including interviewer instructions, interviewer performance, and ultimately whether the respondent knows what the absent researcher means when she answers the question.[5]

---

[4] Question wording and paths through the interview are literally pre-"programmed" when computer-assisted interviewing is employed.

THE SETTING: A family's home with a mother and her two young children, where an interviewer is asking a question from the HOME-SF.

INTERVIEWER: About how many, if any, push or pull toys does [child's name] have?

RESPONDENT: (Pensively, to herself initially) Push or pull. . . . Um . . . I'm not sure what you mean, exactly.

INTERVIEWER: I can't really explain it besides just repeating it and having you think. **It, it's whatever it means to you** [emphasis added].

RESPONDENT: (To herself:)Push or pull toys . . . [inaudible] . . . push or pull . . . [inaudible] . . . maybe four.

3. Interviewer Programming and Standardization. If the researcher is not present, how is his or her intent to be expressed in the survey interaction administered by the proxy — the lay survey interviewer? The usual strategy is to try to program the interviewer's behavior as precisely as possible in an attempt to standardize the survey interaction across interviewers and respondents. Historically, cost concerns and the related need to rely on lay interviewers who are not themselves researchers to collect survey data provided the impetus for the evolution of the preprogramming and standardization that characterize today's typical survey interaction, which in turn can lead to the artificiality, inflexibility, and sometimes lack of clarity evidenced in the example above.[6] Despite such difficulties, the end result of this process is that standardization has become the hallmark of the typical survey interaction. In this context, a primary goal of survey design and interviewer training has been to ensure uniformity in the way interviewers read questions, probe unclear responses, and answer respondents' questions.

---

[5] For the actual 18-month survey, item-by-item definitions were developed (so that interviewers could legitimately provide a definition). However, even this strategy leaves several questions unanswered: how was this issue handled in other studies using the abbreviated HOME Inventory? How might differences in definitions across studies have affected respondents' understanding of the question and their answers? Does such variation affect the validity and reliability of the scales built on such items?

But standardization in itself does not ensure data quality and may actually detract from data quality if some respondents fail to understand the questions, or different respondents attach different meanings to the questions. Thus, the "standardized" stimulus may not be standardized from respondents' perspectives at all. While some within the field of survey research continue to champion absolute standardization (for example, Fowler and Mangione, 1991), others (for example, DeMaio, 1991; Groves, 1987) have expressed concerns about comprehension and constancy of meaning that echo the more vigorous critiques of survey research from outside the field.[7] In fact, some survey interviewers do deviate from the question wording and procedures specified in the questionnaires they administer; some may do so in order to "lubricate" the interaction, particularly in the face of comprehension problems on the part of the respondent, while others simply prefer to substitute their own language for that of the absent researcher.

Indeed, issues of standardization — and departures from standardized procedures and principles — were salient throughout the design and implementation of the New Chance Observational Study. The evolutionary process described by Cannell in footnote 6 was recapitulated in the design of the study, as the laboratory protocols were translated into scripts resembling standardized survey questionnaires. However, it is important to remember that we were unable to provide explicit scripts for every situation that the interviewer would be expected to handle, but instead had to provide interviewers with general principles or decision rules and suggested responses. During data collection, issues of standardization and adherence to procedures were expressed through a tension between two kinds of problems: (1) standardization getting in the way of naturalness of delivery or "appropriate" responses by the interviewer and (2) lapses in standardization of administration of the protocol or application of standard principles in responding to unprogrammed situations. Although presumably accustomed to pairing the establishment of rapport with a professional delivery in the typical interview situation, some interviewers found it difficult to combine naturalness with a standardized delivery in the observational work. It may be that the two types of tasks are sufficiently different that additional training and practice are needed to effect the transition from traditional survey interviewing to the expanded role required in an observational study, especially the ability to respond appropriately to unscripted situations. In addition, there may also be limits to how well lay interviewers can really understand and internalize the *general* principles needed to react appropriately to *specific* situations during the press of the situation, such as the distinction between clarifying task objectives for the mother and providing actual direction as to how she should work with her child. On a more positive note, however, because survey interviewers have little, if any, appreciation of the research hypotheses under consideration or any investment in particular study outcomes, their departures from standardization may be relatively unlikely to involve "experimenter effects" in the classic sense.

---

[6] Cannell (1988) reflects on his own experience in the 1940s doing surveys in which professionals conducted semistructured interviews guided by just a few formal questions, while relying heavily on nondirective follow-up questions. The cost of these interviews turned out to be very high, and it soon became necessary to reformulate them for completion by nonprofessional, less well trained, temporary interviewers. This effort required narrowing the questions and making follow-up questions and probes more explicit. Cost played a role in still another way: the researchers did not have the resources to mine the vast amount of narrative data, so they resorted to simple scales that ignored the richness of the information that had been collected. Since that era, researchers have tended to narrow both their questions *and* response categories at the outset.

[7] Mishler (1986), for example, has questioned whether the structured survey interview can tap the experiences that are important in people's lives at all. Mishler argues for a much more loosely structured approach that requires interviews to be conducted by research professionals, but his position seems more an argument for ethnographic interviewing than a viable alternative within the survey context (other than in pilot or post-survey qualitative work), given the costs involved.

Developmentalists trying to conduct observational research through survey methods or asking the survey community to take on other ambitious assignments may or may not be comforted by the knowledge that survey researchers who do more "straightforward" surveys continually face the same quandaries about standardization that can surface in efforts like the New Chance Observational Study. Traditionally, survey researchers have tried to deflect the issue by writing better and better questions and interviewer instructions rather than by giving interviewers greater discretion to engage in conversation-like activities. The current work followed in that tradition. Efforts to improve questions and instructions are based on evidence that interviewer effects are most likely in situations requiring interviewer discretion, such as in probing open-ended responses, or in asking questions that are challenging for respondents to answer (see Groves and Magilavy, 1986; Groves, 1987). Survey researchers are legitimately concerned that granting interviewers *greater* latitude in their interactions with respondents — even for laudable purposes — will increase the variance associated with interviewers and diminish the absent researcher's control over the interview situation. At the same time, survey researchers are concerned about respondents' comprehension of their questions and ability to formulate accurate answers. As a result, some are proposing to allow interviewers *more* discretion in working with respondents to elicit valid responses. Tensions surrounding the issue of standardization remained an underlying theme of the study and are touched on throughout the remainder of Part II.

B. What the Field of Survey Research Offers in Addition to a Data Collection Capability

A recent review of parenting attitude scales (Holden and Edwards, 1989) within the psychometric tradition of the child development community provides an interesting illustration of the limited communication between the survey research community and those who might commission surveys or wish to work with survey data. The authors are critical of many of the scales created by their colleagues in child development because they find them either unexamined or unsatisfactory with respect to traditional psychometric tools, such as measures of reliability or validity. They take the next step of looking closely at the measures themselves — examining question wording, for example, and similar issues that preoccupy survey researchers — to try to determine what may be wrong with the scales that do not perform satisfactorily from a psychometric perspective. However, the extensive measurement literature emanating from the field of survey research, particularly work on attitude measurement and questionnaire design, plays a relatively minor role in their assessment of measures.

Overall, survey research as a field has spawned a rich literature on the craft during the past 50 years or so. The underlying theme of most of this work is the reduction of error in surveys. Although developmental psychology is turning to survey research primarily for help with problems related to nonobservation (that is, to reap the benefits of larger and more representative samples), it should also be aware of and exploit the survey community's experience in addressing measurement issues, particularly those specific to *survey* measurement.

The process of obtaining maternal self-reports represents the classic survey research model. It is this measurement process, in which a survey interviewer asks a respondent questions and the respondent answers them, that has inspired the bulk of extant methodological research within the survey community. While this body of research is undoubtedly germane to self-reported measures of parenting, it also addresses many aspects of the survey interaction that should help in thinking about the less conventional measurement processes undertaken in this study. In the discussion below we focus first on the design of measurement strategies, chiefly instrument design, then on what is known about the role of the survey interviewer in the measurement process, and finally on a promising framework for conceptualizing and integrating methodological issues in survey research.

1. **Measurement Strategies**. How can the survey literature inform the design of instruments and procedures for survey-based[8] research on parenting? The bulk of the measurement literature has focused on the design of measurement instruments, that is, questionnaires, as opposed to interviewer or mode effects, for example. If this literature yields an overarching lesson, it is a humbling one: that many features of questions and questionnaires can affect survey response.[9] Although the literature does not yield ironclad rules (or even necessarily unequivocal lessons) for designing data collection instruments, it does help in identifying the circumstances under which various threats to measurement may occur. As such, it is helpful both in developing measures and related data collection instruments and in assessing their likely strengths and limitations.

Many issues, informed to varying degrees by the survey literature, need to be considered in developing measures and designing data collection instruments and procedures. For instance, the various response sets (for example, social desirability, acquiescence, extremeness) that may be employed by respondents, and especially individual variation in this regard, have been sources of concern. Work that has tried to validate respondents' answers against objective criteria and had to struggle with defining what constitutes a "match" (for example, Miller and Groves, 1985) raises larger issues as to how closely respondents' answers should be expected to mirror objective "reality." Minimizing errors of recall has long been of particular interest to survey researchers; such work attempts to understand respondents' recall strategies (and the way such strategies may facilitate recall or introduce distortions), individual variation in strategies, and the nature of recall errors, as well as forming a basis for building recall strategies into a questionnaire itself. This work may also help researchers to acknowledge the limits on what can be recalled, for example, the retrospective reporting of prior attitudes. Other work on survey measurement has gone beyond the questionnaire itself and looked at survey design features usually thought to affect only cooperation rates, such as data collection mode or strength of confidentiality assurance, to investigate their possible effects on the substance of the survey response itself.

---

[8] The term "survey-based" encompasses *all* research conducted through the survey model just described, regardless of the specific data collection techniques employed. It is broader than the "interview-based" method that obtains maternal self-reports and interviewer ratings.

[9] Some examples of features that have been shown to be important under certain circumstances are context and order effects; terminology; question style (open or closed); nature of the response scale (number of points, anchors, content); opinion items presented as statements versus questions; use of filters to determine knowledgeability, comprehension, or presence of an opinion; availability of a "don't know" response; the "all that apply" format; and various approaches to obtaining sensitive information.

Andrews (1984) offers compelling arguments for concern with measurement processes and measurement error in surveys — like New Chance — that seek to detect impacts, establish relationships, or predict. Noting the growing recognition of the profound effects that measurement errors can have on statistical relationships — both bivariate and multivariate — he observes that such research endeavors may be particularly sensitive to measurement error. Correlated error affecting multiple measures in similar ways is of particular concern in survey research, because analyses are generally based on multiple measures obtained through the same method, making them subject to method effects. While his arguments support an approach like that of the observational study, with its varied sources and types of measures, they also suggest the importance of making those measures as sound as possible.

2. The Management of Interviewer Effects. What, if anything, can we learn from the survey literature about the ways in which interviewers may affect measurement[10] on a study like this? The need to understand interviewer effects is prompted by a hope that such effects can be controlled when they are undesirable and exploited when they are desirable (Blair, 1980). An example of an undesirable interviewer effect would be the introduction of bias through directive probing of a response. The idea of a desirable interviewer effect may seem counterintuitive. However, if we know that matching interviewers and respondents by ethnicity, for example, leads respondents to overstate the extent of their voting behavior (see Anderson, Silver, and Abramson, 1988a, 1988b), a desirable interviewer effect would be the enhanced validity resulting from the "unmatched" situation. Unfortunately, however, for reasons noted below, less is known about interviewer effects — desirable or undesirable — than about questionnaire design.

As might be expected, the limited work that has been done has examined the interviewer's effect on substantive responses during the traditional survey interview. But unlike the example cited above, most of this work has lacked an independent validating criterion, so that even when interviewer effects are found, their meaning is often unclear. Work on interviewer effects has tended to focus either on the demographic characteristics of the interviewer (or match between interviewer and respondent) or on task demands of the survey interaction, such as characteristics of questions (closed versus open-ended, for example). As observed by Eldred (1994), the processes hypothesized to mediate interviewer effects tend to vary depending on which of these two types of effects is examined. Speculations about effects associated with demographic characteristics of the interviewer have tended to focus on the *respondent's* cognitions and behaviors in relation to perceptions or beliefs about the interviewer. For example, responses thought to be mediated by a "general deference" response set reflect the respondent's sensitivity to the interviewer's group membership and a desire to avoid offending him or her (Reese et al., 1986).

---

[10] Interviewer effects on cooperation rates and item nonresponse have been documented as well: for example, Oksenberg, Coleman, and Cannell (1986) and Singer, Frankel, and Glassman (1983).

In contrast, speculations about interviewer effects deriving from the task demands of the survey interaction have centered on *interviewer* behavior — characteristics of the interviewer-respondent interaction, such as how interviewers probe open-ended or unclear responses (Eldred, 1994). Groves (1987) observes that "the component of variance associated with interviewers is often found to be rather small in professional survey work, but has been found to be larger in responses to open questions dependent on interviewer probing behavior." The mechanisms hypothesized to mediate such effects in less structured situations are rather straightforward (Groves and Magilavy, 1986; Tucker, 1983). Interviewers bring their own value systems, expectations, sensitivities, skill at listening and processing information, and other unique characteristics to the survey interaction. In strictly programmed situations there is relatively little room for these characteristics to come into play. However, when the interviewer must make judgments or exercise discretion (for example, in deciding whether a respondent's answer should be probed or in phrasing a probe), there is a greater opportunity to engrave the mark of the interviewer and, in turn, to exert an unintended influence on the respondent. The presence of interviewer effects in situations requiring more judgment or unprogrammed behavior is clearly relevant to the present work, which required interviewers to make substantive ratings of the home environment and mother-child interaction, obtain time use information through a series of open-ended probes, and administer a scripted observational protocol, while applying general principles to unscripted situations.

Very little research has been done that *directly* assesses interviewer performance, however. Only a few studies have examined deviations from programmed behaviors; interviewer characteristics associated with such deviations; the feasibility of using training, monitoring and supervision to improve performance; and the prediction of field performance from practice interviews (Billiet and Loosveldt, 1988; Blair, 1980; Cannell, Lawson, and Hausser, 1975; Presser and Zhao, 1992; Bradburn and Sudman, 1979). Yet there is increasing interest in the interviewer side of the measurement equation. Fowler and Mangione (1991) urge that even more attention be paid to interviewers in terms of training, supervision, and monitoring than may typically occur. And researchers at the University of Michigan are currently at work on a system of interviewer "behavior coding" that will provide a structured framework for assessing interviewer performance, in addition to informing instrument design (Blixt, 1994).

Given the central role played by survey interviewers in survey data collection, the lack of much information about how interviewers affect measurement may come as a surprise to those outside the field. It is not that survey researchers view interviewers as unimportant, however, but that they rarely have the resources to design and implement studies that make it possible to assess interviewer effects. The relative dearth of work on interviewers is rooted in dual strategies for addressing survey error (Groves, 1987): (1) taking steps to minimize error and/or (2) incorporating mechanisms for measuring error in a survey design. In theory, these two strategies are not mutually exclusive, and, indeed, in a well-designed survey errors of nonobservation (for example, sampling error) are both minimized to the extent possible *and* measured. However, this is generally not the case for measurement error in general, and rarely the case for interviewer effects in particular. Most studies simply cannot be staffed in a way that makes it possible to examine interviewer effects or even to control for them; except in a central telephone survey environment and usually only in connection with methodological studies, cases are not assigned randomly to interviewers and interviewing loads are uneven. The nonrandom assignment of cases, in particular, means that interviewer and respondent effects are confounded in the typical survey. Thus, because of the inability to *assess* measurement error, survey researchers instead emphasize *minimizing* interviewer effects and other measurement error (for example, recall problems).

Since interviewer variance is one of the error components survey practitioners seek to minimize rather than measure, they have traditionally concentrated on strictly programming interviewer behaviors (Blair, 1980) by standardizing instrumentation and procedures (as discussed earlier in this chapter), training, and supervision. Indeed, we followed this tradition in the New Chance Observational Study. However, once survey researchers have made such up-front efforts, most go on to treat survey data as though there are no interviewer effects. Because the number of interviewers on a study is generally small in proportion to the number of respondents, however, interviewers are potentially in a position to exert systematic effects. Thus, some survey researchers take the additional step of limiting the potential impact of any one interviewer by capping the number of cases each is allowed to complete on a given study; others favor letting the "best" interviewers complete as many cases as possible, believing that "positive" interviewer effects are beneficial. Debates about this particular strategy beg the question of how interviewer performance is actually assessed, however. And, as a practical matter, such methodological debates may occur far from the ears of those responsible for staffing decisions and day-to-day survey management.

With the expanded responsibilities assumed by interviewers on a study like this, there is reason for attention to the interviewer's role. But without much empirical information to draw on, researchers at least need to be sensible about what they ask survey interviewers to do and to consider the selection and credentialing processes, training approaches, and supervisory structures needed to support innovative data collection efforts. These issues are explored in the chapters that follow, which report on the implementation of the New Chance Observational Study, make recommendations for future work, and, it is hoped, shed light on what it is "sensible" to ask interviewers to do.

3. **A Promising Framework for Conceptualizing Challenging Survey Efforts and Integrating Research on Survey Measurement.** Although the field of survey research has had a rich and longstanding tradition of research on the recall of information during the survey interaction, the past decade or so has seen the emergence of a broader interest in the cognitive aspects of survey response. This has come about through the collaboration of cognitive psychologists and survey practitioners working together to improve the quality of survey response. The cognitive movement goes beyond recall to examine the myriad cognitive demands of survey response systematically and to conceptualize the processes involved in answering survey questions (see Tanur, 1992). By breaking the survey interaction down into components, based on an information processing model[11] (Cannell, Miller, and Oksenberg, 1981; Dippo and Norwood,

---

[11] In this model, survey response has four distinct stages: comprehension (of the question), retrieval (of pertinent information), judgment (in formulating an answer that is responsive to the question), and communication (of the response, within social and other constraints of the interview situation). While most applications of this model have been to respondents, it has also been extended to interviewers, thereby creating the potential for a unified approach to survey measurement.

1992; Turner, Lessler, and Gfroerer, 1992), the cognitive approach provides a conceptual framework for thinking about the demands of the survey interaction on both respondents and interviewers. It also provides a framework for integrating what is known about the instruments used to collect data and the interviewers who administer them. We will turn to this approach in greater detail in Chapter 13 in reflecting on the implementation and results of the New Chance Observational Study and the lessons they hold for similar endeavors.

C. What We Can Conclude About the Survey Method

Despite the sheer volume of findings related to survey measurement, only rarely does the research literature point unequivocally to a specific survey measurement strategy for a particular survey. Most of the research, even that involving experimental manipulation of question wording or procedures, for example, reports on variations specific to a particular context. Over time, many general principles have been distilled from this work, but survey designers still have to make judgments as to how these principles apply in specific contexts that may differ from those that originally generated the methodological findings. Thus, the literature does not offer pat recipes for instrument design, although it does serve to raise the consciousness of those designing surveys and suggest "worry lists" of things to be concerned about in the course of design.

Nonetheless, it is rare for a survey design to incorporate all that is known or believed to be optimal: while inherently incompatible priorities may get in the way, resources and feasibility are more often the problem. For instance, because surveys typically address many topics, measurement of any given construct may suffer because the number of items that can be devoted to it must be limited. As another example, because resources are finite, pretesting cannot go on indefinitely even if additional pretesting is thought to be beneficial. Finally, "slippage" can and does occur in the actual implementation of a survey.

Contemporary survey research always involves compromise, as Groves (1987) observed in considering why methodological wisdom is not always incorporated into surveys: "To become perfect measuring devices they must stop being surveys (as we know them)." As developmental psychology looks to the survey community to leverage larger and sounder samples, it will be important for developmentalists both to recognize the constraints of the survey method *and* to develop measures and measurement processes that build on its accumulated body of wisdom.

III. An Overview of the Methodological Issues in Part II

The overall theme of Part II is the measurement of parenting constructs and child outcomes through the multifaceted approach taken in the New Chance Observational Study. The next three chapters focus on the measurement processes underlying the parenting and child outcomes measures discussed in Part I.

This study affords a unique opportunity for examining methodological issues in the measurement of parenting and child outcomes for two reasons. First, as previously discussed, there is the challenge involved in exploiting the survey model in reasonably innovative ways. Second, there is the diversity of the measures and data sources: taped mother-child interactions coded under rigorous conditions in university laboratories, self-reports of the mothers, and ratings by survey interviewers of mother-child interactions and the home environment. In terms of content, both affective and cognitive domains are well represented, along with other topics such as time use, especially time spent by mothers with their children in various activities, and subjective maternal reactions to parenting. Finally, measures that have already enjoyed considerable use are included, along with new measures developed for this study with the hope of improving the measurement of parenting constructs.

Table 10.1 presents a typology of the parenting measures used in this study classified according to the measurement processes that resulted in them and the roles of those involved in producing the data. The framework reflects the major themes discussed so far and foreshadows the discussion in the remainder of Part II about the strengths and vulnerabilities associated with various measurement strategies.

Table 10.1

Structural Typology of Measures
in the New Chance Observational Study

Type of Measure:	**Maternal Self-reports on Parenting and Time Use**	**Interviewer Ratings of Mother-Child Interaction and the Home**	**Expert Coding of Structured, Videotaped Mother-Child Interaction**
What Role Does Mother Play?	Answers questions asked by interviewer	Allows interviewer into her home for interview or observational session	Is videotaped working with her child on a structured set of tasks
What Role Does Interviewer Play?	Asks mother questions and records her answers	Observes home and mother-child interaction and records her own observations and ratings	Administers a structured interactive protocol to mother and child
Who Is the Ultimate Informant?	Mother	Interviewer	Expert coder in university laboratory
Who Else Is Key to the Measurement Process?			Videographers who record the interaction
	Absent researchers who design the data collection instruments and procedures and who are "represented" during the visit by the survey interviewer		
	A survey operations infrastructure responsible for oversight of the data collection and supervision of interviewers		

In the table, the specific roles of mother, interviewer, and others in generating the measures are delineated, and one individual is designated "informant" for each measure. "Informant" is really the proximate informant, since, as suggested by the table, there is always more than one party involved in producing each measure.[12] For the measures in the first two columns, "informant" assumes the conventional meaning of the person who answers the questions or completes the checklist. In the third column, the expert coder, who completes ratings based on the behavioral interaction, is designated as informant. Viewing the framework another way, the first column represents the traditional survey model, in which an interviewer asks a respondent questions and records answers, while the measures in the second and third columns expand upon that role — first by making interviewers a direct source of substantive data and then by making them responsible for administration of an observational protocol.

As a complement to the substantive findings of the New Chance Observational Study presented in Part I, Part II views the research from a methodological perspective. Within the context of the issues laid out in this chapter, it discusses the implementation of the research. Following in the tradition of the survey model discussed above, implementation of the research focused on efforts to minimize measurement error rather than a study design that would facilitate studying or measuring various sources of error. Thus, the methodological discussion that follows by no means provides answers to all of the questions that may have been raised by the discussion of the measurement process earlier in this chapter. While we have some data to help in assessing interviewer performance, for example, we are not in a strong position to tease out the effects of their performance on the behavior of the dyads who participated in the study. Still, we see value in at least defining the key issues, since doing so begins to open the "black box" of measurement if only to generate an inventory of issues to address in survey design.

Chapter 11 describes how the study was implemented — from translating the laboratory-based protocols for survey administration to training field staff. Thus, it explains how the team sought to address measurement issues like those raised throughout this chapter in the study design and survey management. The discussion attempts to provide sufficient detail about implementing the study to make clear what is involved in doing such research within the framework of a survey and to provide a context for subsequent discussion of lessons learned from the effort. Chapter 12 describes the observational sessions completed during this study, in terms of various situational complexities associated with completing these sessions in the homes of low-income families. This material is aimed at helping readers understand both the challenges facing the field staff and the potential for variation in the data collection setting from one session to another. Chapter 12 also considers various indicators of the dyads' subjective experience of the session. Finally, with the data that were available (despite the study not being designed to investigate interviewer performance or effects), it attempts to assess the performance of the lay survey interviewers who collected the data. Chapter 13 provides an evaluation of the overall effort, with specific suggestions about what could be done differently in future work. Chapter 14 concludes Part II with an assessment of the measurement implications of the New Chance Observational Study.

It should go without saying that Part II does *not* describe a model effort with unlimited resources that crafted a "seamless" integration of the substantive concerns of developmental psychology with the methods of survey research. That kind of effort is rare in contracted survey research, and a project undertaken by a consortium, most of whose members had not previously collaborated, would seem an unlikely candidate for breaking such ground. Rather, the next few chapters represent an opportunity to step back and reflect. For developmentalists who wish to take advantage of the samples available through survey research, it is hoped that the discussion will increase awareness of the complexities inherent in the survey measurement process itself. For survey practitioners, it is hoped that this material will inform the design of studies with some of the demands of this one. For researchers interested in collecting data on parenting and child outcomes, Part II is designed to help in assessing the strengths and weaknesses of alternative approaches.

---

[12] Thus, one thing to keep in mind is the potential influence of the interviewer even when she (all were female) was not the proximate informant for a given measure (for instance when obtaining maternal self-reports and particularly when administering the observational protocol).

Chapter 11

Implementing Observational Research Within a Survey Context

Carolyn A. Eldred

This chapter describes how the New Chance Observational Study was implemented, including design of data collection instruments and procedures and the data collection process itself. It demonstrates how the issues discussed in the previous chapter were addressed in practice. The discussion is intended to provide both sufficient concrete information to inform other similar efforts and a context for the methodological assessment that follows.

In the discussion of design, the chapter addresses the strategies that were employed for adapting laboratory-based protocols for survey administration, including strictly scripting presentation of the protocol and providing guidelines for handling the situations that could not be explicitly scripted (because their number was potentially limitless). This chapter also discusses design and pretesting of the data collection instrument and the design of training materials. As part of the discussion of data collection, the chapter addresses the organization and management of the field effort, including the challenges it presented. Also addressed are mechanisms employed to enhance data quality, including training of interviewers and videographers and ongoing quality control mechanisms.

Implementing the observational research that is the central subject of this monograph involved both the design of data collection instruments and procedures and the data collection effort itself. In carrying out these tasks, we confronted and sought to address the host of survey measurement issues discussed in the previous chapter, most notably the overarching issue of adherence to the goals and objectives of the "absent researchers." In this chapter we attempt to provide sufficiently detailed documentation of how we approached this work to inform similar efforts by other researchers and to provide a context for the methodological assessment in the next chapter and the recommendations for future work in the succeeding chapter.

I. Instrument and Procedures Design

A. Strategies for Adapting Laboratory Protocols for Survey Administration

Placing responsibility for protocol administration in the hands of survey interviewers had a number of implications. Since the interviewers lacked the theoretical underpinnings to conduct the sessions while working only from a loose outline, the strategy for enabling the "absent researchers" to maintain control over the session was to program interviewer behavior carefully.

This meant that interviewers required a data collection instrument resembling a survey questionnaire, which specified the precise language to be used and the actions to be taken.[1]

Programming interviewer behaviors in this way was considered essential, even if it meant that considerable practice would be needed before interviewers could administer the protocol in a reasonably natural way. It was not an option to ask interviewers to memorize the protocol so that they could present the tasks without referring to the script. This would contradict a basic tenet of survey interviewing — to *read* each question actually as written no matter how well the interviewer knows the questionnaire. More important, interviewers typically work on several surveys simultaneously, and their workload on the observational study was probably lighter than on most of the other assignments they would have had at the same time. This suggested the strong possibility of interference from other material even if interviewers successfully memorized the protocol at the outset. Clearly, survey interviewers would require considerable structure if they were to act as the agents of the researchers in this endeavor. Thus, the data collection instrument developed for their use embodied the kind of structure usually found in a survey questionnaire.

Despite the interviewer programming built into the instrument and the use, wherever possible, of questionnaire formatting conventions that would be familiar to interviewers, the data collection instrument differed from a survey questionnaire in significant ways. Most notably, it was not a plan for asking questions and recording answers. Rather it was a script containing the words to be spoken to the mother in explaining the tasks and in administering them to the mother and child together. It also included detailed instructions for presenting, arranging, and withdrawing the various props (book, games, and gift) used in the tasks and for coordinating management of the props with the script. Also included were language and instructions for verifying that the mothers understood the objectives of the task, decision rules for determining how much time to spend on each, and, for one task, decision rules as to whether a simple or complex version, or both versions, of the task was to be administered. The interviewer's role was to set the stage and allow the mother-child interactions to occur. Except for some ratings made by the interviewers, the actual data from the observational segment of the session were collected passively — by audiotape and videotape — rather than through the active involvement of the interviewer.

A survey questionnaire typically tries to anticipate and provide for every scenario that can be tapped by a question; this is generally done by creating response categories for recording answers to a question that are exhaustive and (usually) mutually exclusive. Because the questionnaire designer dictates the categories that will be used, it is generally not very difficult to construct a set of categories that must logically encompass all situations. A second way that questionnaire designers provide for the variety of real-world situations that may exist is by employing "skip patterns," in which questions are asked or not depending on the answers to prior questions or characteristics of the respondent. Like response categories, skip patterns must be designed to anticipate all of the situations to be encountered. The data collection instrument for the observational study did not need to follow in the survey tradition of anticipating every possible response, however, because the path through the observational script was not dependent on the responses of the dyad and, by and large, did not require the interviewer to intervene in the activities. More important, it *could not* anticipate the virtually infinite range of behaviors and events that might occur during the session, not only involving the mother and child but also other people and events in the household. Questionnaire designers artificially constrain respondents' answers by generally using "precoded" response categories. The observational session, by contrast, provided far less constraint than even the most open open-ended question in a survey questionnaire. If administered properly, it was intended to provide mother and child with a blank slate on which to engrave their personal response to the activities they were asked to carry out.

---

[1] The data collection instrument developed for this study and the Interviewer's Training Manual are available upon request from MDRC.

Since it would have been impractical to embed instructions for responding to even a sampling of hypothetical situations in the data collection instrument without rendering it unusable, a separate set of guiding principles, rules, and examples was developed. Interviewers would need to internalize this material and be able to apply it in the actual data collection situation. They were not expected to make their own judgments as to how to handle situations that would arise. Still, even if they were very familiar with the guidelines, they would need to label a situation correctly in order to respond appropriately. For instance, they needed to understand the distinction between a mother's request for clarification of the instructions or objectives of a task and a request for guidance as to how to work with her child, since these two situations called for quite different responses. Thus, an integral part of the study instrumentation was the interviewer's training manual — particularly those sections that tried to give interviewers tools to help them handle situations that could not be scripted.

B. Instrument Design

As noted in Chapter 2, each observational session began with a book reading activity and discussion developed by the Snow laboratory at Harvard University. It then presented four teaching tasks developed by Harrington, Block, and Block (1978) and used in research conducted in the Egeland laboratory at the University of Minnesota: model replication with blocks, naming objects with wheels, sorting simultaneously on either one or three dimensions, and navigating a maze on an Etch-a-Sketch toy. The final activity was presentation of a gift.[2] Using a script embedded in the questionnaire, the interviewer was first expected to explain all of the tasks to the mother while the child was distracted by the videographer or others in the household if possible. The child then joined the mother and interviewer. In administering the tasks, the interviewer continued to follow the script, reminding the mother what each task required while presenting the props for the task. Throughout the session, the interviewer was to interact primarily with the mother rather than the child and to deflect requests for advice or direction.

The initial design of the instrument involved making tentative decisions about how to mesh the tasks from the two laboratories so that they were presented in an optimal order and formed a coherent whole. This required attention not only to the order in which the tasks were presented, but also consideration of the seating arrangements and work space appropriate for each. At the same time, steps were taken to create an instrument that functioned in many ways like a survey questionnaire. The protocols that had proved workable when administered by researchers or their assistants were translated into much more explicit instructions. If the interviewer was required to do or say something, it was reflected in the instrument. In the process, the team was forced to think and make decisions about exactly what would happen during the session, the precise words the interviewer would say, and any contingencies or decision rules that would have to be built in.

---

[2] The Minnesota laboratory would later apply its coding scheme to the session as a whole; the Harvard laboratory would code the book reading and discussion activities, as well as the Minnesota wheels task.

Particularly once pretesting was under way, the team also had to think about operational issues that might not arise in the laboratory or even necessarily in university-based studies involving home visits. Issues such as how to manage the bulk and weight of the interviewer's props, how to arrange adequate lighting for videotaping, how to conduct the session if no table were available, and what to do about interruptions from other children were addressed by the team and led to either modifications of the original instrument and procedures or additional specification. Pretesting also provided experience germane to such substantive issues as the appropriateness of the tasks for the age range and level of disadvantage represented by the New Chance sample, the dyads' engagement in the activities, and the degree to which the mothers and children seemed to exhibit reasonably natural behavior without becoming unduly anxious about their performance.

C. Pretests of the Data Collection Instrument and Procedures

Because the tasks to be administered had previously been used extensively, full-scale piloting was not undertaken. Rather, two small pretests were conducted to help in refining the "translation" of laboratory protocols for survey administration, check on the effectiveness of the tasks among young disadvantaged respondents, and investigate a number of practical implementation issues.[3]

Even these small pretests anticipated a variety of situations and issues that would arise again in the main study and that are documented in Chapter 12, which presents findings on administration of the observational sessions. For instance, the home environments in which the data collection occurred were typically characterized by a host of distractions and interruptions from television and radio, children, and other sources. A kitchen or dining table was not always available for work space: in a few cases the dyad worked on the floor and in another case used a bed.

The pretests revealed how the interviewers' personal styles could affect the way in which the sessions were administered. Some interviewers seemed to have particular difficulty distinguishing situations warranting their intervention from those that did not. Instead of sitting back and letting the behavior of mother and child unfold on its own, they would intervene, taking on the more activist, controlling role of questionnaire administrator. These findings underscored the need to program interviewer behavior as precisely as possible and to train interviewers to handle situations that could not be precisely programmed.

---

[3] The two pretests were conducted in December 1991. The experience of each pretest was evaluated by the team and revisions made as a result. Nine sessions were conducted in all: the first pretest included four young mothers participating in an educational program for teen mothers at Temple University in Philadelphia, while the second included five New Chance sample members (three experimentals and two controls) from the Allentown, Pennsylvania, site, which would not be involved in the actual observational study. The mothers in the two pretests ranged in age from 17 to 23 at the time of the session; all but two were 19 or 20. Like the target sample for the observational study, their children ranged in age from 30 to 60 months; six were between 38 and 52 months. Six of the women were African American, and three were white.

Making sure the mother understood each task first arose as a concern during the pretesting and remained an issue during the data collection, since interviewers were charged with the somewhat conflicting goals of clarifying the objectives of the task if the mother failed to understand the task, without making suggestions or giving examples of how to do it. (This turned out to be primarily a problem on the wheels task.) Similarly, merging tasks that had not previously been used together but that required different seating arrangements (the book reading activity and teaching tasks) created a situation that was difficult for interviewers to manage (without becoming too intrusive) and that often resulted in less than optimal arrangements for some activities. This remained an issue throughout the study.

Requirements that interviewers make subtle distinctions — reinforce effort rather than performance, for example — posed problems during the pretests and continued to do so during the study. Potentially conflicting requirements also surfaced as an issue and remained so throughout. For instance, interviewers were instructed to deliver the initial instructions to the mother without the child present and were given a number of procedures to employ to keep the child from hearing about the tasks from the interviewer rather than from the mother later. At the same time they were told not to "banish" the child if this upset him or her too much. Finally, there were issues that did not elicit consensus within the team. The issue of who should be responsible for "crowd control" in the household — the interviewer or the mother — was resolved in favor of the mother,[4] so as not to undermine her authority, but some team members continued to feel that a firm word from the interviewer would carry more weight and improve the quality of the session. Another issue that seemed to elicit a range of reactions from team members during the pretests and throughout the study concerned the degree of interest or involvement that the interviewer should display while the dyad worked on the tasks. In the end, interviewers were instructed to display "polite interest," neither ignoring the dyad's interaction nor becoming involved in it, but for some this was a difficult balance to achieve.

Many other issues were successfully resolved through pretesting and, by and large, did not resurface during the study itself. Some of the tasks worked well in the field from the very beginning, while others required refinement. For instance, the book reading activity, maze task (Etch-a-Sketch), and gift (kaleidoscope) all worked well and received only minimal revision as a result of pretesting; in particular, the book-reading activity was effective as the initial icebreaker. The wheels and sorting tasks posed some difficulty for the youngest children and were modified as a result of the pretests: the revised procedures provided for moving on to the next task if the child was unable to name any objects with wheels, and an additional, simpler, sorting task was developed for the youngest children. Finally, pretesting suggested modifications in the number, size, and color of the blocks used in the modeling task. In addition, the pretests helped in refining the language used to present the tasks, especially in terms of striking a balance between motivating the mothers to take the tasks seriously and minimizing the perception that they or their children were being evaluated.

Technical refinements were also made based on pretesting. It was decided to use an auxiliary light directed at the ceiling as a fairly unobtrusive enhancement to the overall light level. An extra microphone was located close to the work area to enhance the sound beyond that avail-

---

[4] By the researchers, but not necessarily by the interviewers.

BEST COPY AVAILABLE

able from the video recorder. To aid the Harvard team in its transcription, a separate audio recording was also made. The pretesting also emphasized the need to be sensitive to problems such as backlighting that could make the participants' faces hard to see, noise from inside or outside the home that could be picked up clearly on the tape, and props being placed or dropped on the auxiliary microphone. Finally, it was decided to use a large square of pale yellow felt to cover the work surface. Originally introduced to make it possible for the dyad to work on the floor if necessary, it came to enjoy routine use because it met several needs: it protected the mother's furniture, it provided a background that highlighted the props on the videotape, and it blunted the noise from the blocks and other props on the work surface.

Another decision was made *in spite of* the pretest experience. As an incentive for participating, mothers were given a choice of receiving either a check for $15 or a copy of the videotape of their session. All but one of the mothers chose the videotape. Despite this preference, the team determined that making the videotape available as an incentive posed insurmountable problems. A key issue was reactivity. In what is already an unnatural situation, the objective is for the mother and child to focus on the tasks rather than the camera and to let their behavior unfold naturally. The session should not elicit a "performance" to be shown later to family and friends. There was some evidence in the pretests that this was happening; some of the mothers dressed up and a few had the session administered at their mothers' homes, which may have been more attractive their own. Just as important were the potential difficulties in maintaining the confidentiality of the tapes. Clerical errors, for example, could easily result in tapes being switched and sent to the wrong participant. Delivery problems at the receiving end could pose further risks of a breach of confidentiality. On the other hand, the mothers in the study would have recently had their 18-month New Chance interview and received a monetary incentive. Since this would be their "default" expectation, the team decided that it would not be necessary to offer the videotape.

D. The Design of Training Materials for Interviewers and Videographers

The pretest experience was useful not only in refining the study procedures but in suggesting the kinds of information and guidance that interviewers and videographers would need to do their jobs well. The basic resources consisted of a procedures manual for the study and a training videotape. Interviewers for the observational study also worked on the New Chance 18-month survey, for which they had already received training (using videotaped sequences and still photographs) to make the HOME-SF and related ratings also called for by the observational study.

The procedures manual developed for the study was similar in many ways to manuals used for more typical surveys, covering a mix of substantive and administrative requirements and emphasizing use of the data collection instrument. The manual also addressed topics specific to the observational study such as the ways in which the assignment resembles and differs from traditional survey interviewing, special requirements for conducting the observational session, principles for relating to the child during the visit, definitions of questionnaire items not covered in training materials for the 18-month survey, detailed guidelines for handling unscripted situations, coordination in the field between the two members of the data collection team (interviewer and videographer), and technical instructions related to the taping. The manual was designed not only for training the data collection teams but also to serve as a reference resource throughout the study.

The manual tried to help interviewers understand which interviewing principles they would have to leave behind when conducting the observational work and which had parallels in the observational session. The key characteristic of the observational work is the need to retreat and temporarily suspend responsibility while the mother-child interaction is under way, rather than orchestrating everything that occurs: the interviewer is a facilitator rather than a participant. Nonetheless, as in a traditional survey, the interviewer is expected to make the respondent feel comfortable and to maintain a neutral, nonjudgmental stance; she does not indicate her own opinions or give any reaction to the opinions or information from the respondent. In both situations, the interviewer is to refrain from influencing the respondent. In a traditional survey, for example, an interviewer probes an unclear response neutrally ("I'm not sure I understand — could you tell me a little more about that?") rather than suggesting what the respondent might have meant and asking for confirmation. A traditional survey also requires an interviewer to deal with situations that are not explicitly scripted. For instance, if a respondent asks for clarification of the meaning of a question, the interviewer must know what she is expected to say, which is generally specified in question definitions in the procedures manual. Similarly, if the respondent gives a response that does not adequately answer the question that was asked, the interviewer must know what to do to elicit an adequate answer. These contingencies have parallels in the guidelines specifically developed for keeping the observational session on track, and it was hoped that interviewers would be able to recognize the familiar principles and act accordingly.

Interviewers were given five principles that together provided a framework for their work:

- Study procedures must be standardized.
- An unnatural situation should be made to feel as natural as possible.
- The interviewer and videographer must not influence the mother-child interaction.
- The mother must be thoroughly familiar with the tasks *before* attempting to explain them to her child.
- The interviewer must never offer suggestions about how the mother should work with her child.

Interviewers were also provided suggested responses to questions about such matters as the purpose of the study, requests for guidance or feedback, whether the tasks are timed, and whether the families may keep the toys (props). In addition, they were given guidance about how to keep the child from listening to the interviewer's initial explanation to the mother and how to deal with noise, distractions, and interruptions in the immediate environment. Finally, they were given approaches to follow and language to use in handling a host of highly specific situations, for example, in the event that the child wanders off, the child is very close to success when the time limit is reached, or a younger child in the home needs attention. Because a key objective was to help interviewers learn when *not* to intervene, examples of situations in which they were to do nothing were also provided. These were situations that interviewers could be tempted to identify as problems — and, as problems requiring their intervention, for example:

- The child is not following the "rules" of a task (for example, he is building a house instead of replicating the model on the block building task), and the mother is not reminding the child or explaining the task.
- The child needs help with a task (for example, cannot reach the blocks she wants) and is becoming frustrated, but the mother does nothing.
- The mother does the task for the child completely (for example, completes the sorting board without involving the child).

A second key resource for training interviewers and videographers was a videotaped observational session. The session was not taken from the pretests but was completed with a dyad recruited specifically for this purpose after the procedures had been finalized. As intended, it depicted a generally exemplary performance; trainers took advantage of the few errors on the tape to emphasize the principles violated by the data collection team and the importance of adhering to study procedures.

II. Data Collection

A. The Organization and Management of the Field Effort

This kind of study was new not only to the field staff but also to those managing the effort. Supervision of the interviewers tended to follow established procedures, despite quality control needs that distinguished the observational study from those of the usual survey and coordination problems that developed in the field. One impediment to developing a dedicated survey management structure custom-tailored to the requirements of this study was the fact that it was actually "embedded" in the larger New Chance survey effort. Once the observational study was under way, the same interviewer generally conducted both the 18-month interview and the observational session (typically receiving both assignments at the same time), and interviewer assignments and production reporting for the two studies tended to be intertwined. An important positive step, however, was the assignment of a single interviewer supervisor to handle substantive supervision for the observational study, instead of relying on the dispersed network of local supervisors who were overseeing the 18-month interviews. This was valuable not only because this supervisor was highly competent and understood the study well, but also because the relatively small number of cases in each site, distributed over a fairly lengthy field period, would have made it harder for six or seven separate supervisors to remain proficient throughout the course of the study.

Inclusion of videographers and the need to conduct the field work in teams created additional challenges. Survey research organizations are accustomed to managing interviewers who have learned the conventions of the survey world and of the particular organization. While most interviewers do not work for one organization exclusively, there is generally some continuity in their relationship with the organizations for which they interview. They have probably worked on previous studies, may be working on other current studies, and probably expect to interview for the organization again in the future. By contrast, the videographers were recruited specifically for the observational study, with no expectations of an ongoing relationship. The fact that they were not interviewers but had a technical role, however straightforward, created some confusion about how they should be managed and suggested the possibility of a technical chain of command outside the usual survey management supervisory structure. In fact, such a parallel supervisory structure evolved during a videographer recruitment process that worked through university film or communications programs and designated a faculty member as "location coordinator" to coordinate local videographers and equipment. As a result, for almost the first half of the study, the technical side of the work was not well integrated into survey management. Later, location coordinators were eliminated, some restaffing of the videographer force occurred, and intense efforts were undertaken to improve communication.

The use of two-person teams posed additional complications. As in some other studies requiring the use of teams, the logistics were difficult and interviewers were frequently disgruntled. Particularly because this was an economically disadvantaged sample, with ongoing tracing required to locate respondents and some sample members lacking telephones, the need to coordinate with a second team member made the work harder. The fact that many of the videographers were students — and all were strangers to the survey field — contributed to the difficulty of establishing viable working partnerships and fueled a perception on the part of some interviewers that the videographers were insufficiently committed to the work. Maintaining a balance between having enough videographers at each site and providing each with enough work to lead to a commitment to the study was another challenge, given the relatively small number of cases in each site to be completed each month but a need for considerable flexibility in scheduling sessions. This problem was exacerbated by the fact that the interviewers were responsible for lining up a videographer at the time they made an appointment with a respondent (central office involvement in this process was rejected after some consideration) and by the need for videographers in most sites to share equipment. The consolidated survey management structure that characterized the second half of the study, along with some restaffing in the field, made these challenges somewhat more manageable as the work went on. Yet because of these objective challenges this remained a difficult field effort.

B. Mechanisms to Enhance Data Quality

1. Training Data Collection Staff. While the first step toward data quality is the sound design of data collection instruments and procedures, the second step is training data collection staff to use these instruments. The procedures manual and sample tape developed to train interviewers and videographers have previously been described. This section discusses how the training was conducted.

The first component of the training consisted of completion of a home-study package. Trainees were expected to study the procedures manual at home and review the sample videotape. In addition, interviewers were expected to practice delivering the observational script and to complete a quiz covering the major requirements of the study.

A one-day training session followed. A joint session of interviewers and videographers occurred in the morning. This session began with an overview of the project, followed by a critical viewing of the training videotape, with discussion of the points that it illustrated. In the next module, interviewers and videographers practiced working together in assembling and removing the various props. The pretest work had revealed the complexity and potential awkwardness of manipulating the props and suggested the desirability of having the videographer help the interviewer. This exercise allowed two-person teams to work on handling the props smoothly and efficiently before interviewers had to overlay the script delivery (and reaction to unscripted situations). The joint session concluded with a discussion by the group of the need and mechanisms for close communication between interviewers and videographers.

In a separate session in the afternoon, videographers were introduced to the specific equipment to be used in the study and received training on their technical responsibilities. The afternoon session for interviewers began with review of the maternal self-report items used in the observational study questionnaire but not in the 18-month survey, which would be completed during the interview segment of the observational session. This was followed by a review of the quizzes completed as part of home study. Then, with trainers playing the parts of mothers and children, each interviewer administered two mock observational sessions that had been scripted in advance to include a variety of situations for them to handle. The final component of the "classroom" training covered procedures for scheduling the observational session, especially handling cases who were about to age in or age out of eligibility for the study.

As the last step in the training, each interviewer completed a practice observational session with a family in the community similar to those in the New Chance sample. Interviewers and videographers received feedback after review of their work by members of the observational study team before beginning work with actual sample members.

As previously noted, material that interviewers had been required to master for the New Chance 18-month survey was not addressed in the observational study training. This included instruction on making HOME-SF and other ratings tapping parenting constructs. Since measures deriving from these ratings are among those discussed elsewhere in this monograph, a brief discussion of how interviewers were trained to make them is germane to this methodological work.

Interviewer ratings were to reflect both observations of the physical environment of the home and observations of maternal and child behaviors or interactions. Different sets of stimulus materials were prepared to deal with each. For assessing the physical environment, a picture book containing 24 still photographs was compiled. These photographs were taken in the homes of low-income families in locations across the country, many in highly disadvantaged urban neighborhoods. A larger pool of photographs was initially selected for rating by project staff familiar with the HOME-SF, along the four dimensions tapped by the HOME-SF environmental items (safety, cleanliness, clutter, brightness/monotonousness). The final set of photographs was selected to provide a range of examples for each dimension.

For the behavioral assessments, a videotape was made consisting of excerpts from pretest interviews for the New Chance 18-month survey. Some of these pretest interviews were recorded with the explicit purpose of assembling footage for this training tape, and respondents consented to such a use of the taped material. Vignettes were selected to provide a range of interactive styles.

Prior to interviewer training, two members of the team independently made all of the environmental and behavioral ratings and discussed areas of disagreement (which were reasonably infrequent). Often, difficulties in making or agreeing on the ratings reflected the limitations of the available information (for example, seeing only part of one room, having a sample of behavior only a few minutes long, viewing a situation that was clearly temporary). Limits on the amount of information that could be conveyed by these materials also placed limits on how the materials could be used in training. In particular, it did not seem feasible to hold interviewers to an absolute standard of "right" answers, especially for ambiguous cases. Instead, the materials were viewed as a teaching tool, with discussion intended to clarify the meaning of the rating items. For training, interviewers were required to make their ratings independently before coming to the training conference. To highlight cases and items warranting special attention during training, their ratings were compared with those previously completed by team members. Tapes were viewed again as necessary and the rationale for the ratings discussed. During home-study the last few ratings were to have been made by leaving a brief interval between watching the tape and recording the ratings. This effort to build in the interference of the real world was taken a step further during training, by requiring the interviewers to engage in another activity (recording the respondent's answers to the interviewer's questions on the videotape) while at the same time trying to keep an observant eye and ear out for information germane to the ratings.

2. Quality Control. Ongoing quality control was perhaps even more difficult to integrate into the survey management process than the supervision of videographers. Typically, for survey efforts conducted in person, the interviewer's performance is evaluated indirectly based on the paper questionnaire or computer record of the interview. In the course of routine coding and editing this output can be reviewed and any problems fed back to the interviewer. Centralized telephone interviewing permits evaluation of actual performance, but it is done in real time in the course of supervision on the interviewing shift by listening in on segments of each interviewer's work. Post-survey validation contacts with respondents also typically occur, at minimum to confirm that the interview occurred and that the interviewer performed professionally, but also sometimes as a more comprehensive check on data quality. In contrast to these models, this observational work resulted in videotapes lasting at least 30 minutes; quality control required viewing each of these tapes individually. Further, there would have to be a feedback mechanism to communicate information in a form that the data collection teams could use to improve performance. But in the context of routine survey operations, there was no natural "home" for the dual functions of reviewing videotapes and feeding back interviewer evaluations, nor even an obvious way to combine these two highly related functions. Nor was there a systematic mechanism for communicating sufficiently specific feedback for the interviewers (and videographers) to use, particularly with the passage of time before such feedback could be delivered and the inadvisability of returning tapes to interviewers.

The first 35 tapes were carefully reviewed by team members and a critique prepared of each tape. In the course of these individual reviews, two detailed field memos were prepared and disseminated to all field staff. These memos communicated only a limited amount of new information, such as a change in how the discussion following the book reading was to be handled or clarification as to how a contingency should be handled. Mostly the memos repackaged and reemphasized information that had previously been presented in the procedures manual and during training. The specifics of the individual critiques might actually have been a more effective vehicle for communicating these lessons to individual interviewers, and the interviewer supervisor did attempt to provide this feedback. But communication was hampered by the fact that the supervisor had not viewed the tapes herself, some time had passed since completion of the session, and the interviewer had to rely on memory to try to reconstruct what she had done.

Because of resource constraints, it was necessary for senior team members to delegate responsibility for quality control after review of the initial set of tapes. Ideally, this function would have been performed by someone with direct responsibility for the interviewers, so that quality control functions could have been integrated into the ongoing interviewer supervision and survey management process. However, the only person actually budgeted to watch every tape was the head videographer, who reproduced copies for the university laboratories. Since he was reproducing the tapes in real time and was responsible for monitoring technical quality, watching the tapes was to a large degree already part of his work. The tapes reviewed by the team and the critiques that had been prepared were used as a training tool to teach him how to identify problems warranting feedback to interviewers. While he was not trained as a researcher, neither were the interviewers nor those in the supervisory structure. Thus, the study design had already embraced an assumption that nonresearchers could understand and apply the principles underlying the data collection task. While this individual successfully performed this function, the function itself did not become well integrated into the interviewer supervision process, which was largely focused on production and on the logistical difficulties in the field.

Despite the fact that quality control efforts did not conform to an ideal model, they did form an important basis for the individual and group feedback and for selective restaffing of the data collection effort. In addition, the results of these attempts have already given rise to an evaluation system for systematically reviewing tapes in the future (discussed in Chapter 13), as well as a vision for integrating quality control into ongoing survey management, rather than viewing it as the province of specialists. These efforts have also yielded anecdotal evidence of differences in the ability and/or willingness of interviewers to adhere to an observational protocol, tailor their behavior to study requirements, and benefit from subsequent feedback. Thus, a fourth tool in addition to sound instrument design, thorough training, and ongoing quality control may be needed; this tool is a credentialing process that selects interviewers carefully and requires them to meet an established performance criterion before work can begin.

In Chapter 12, we assess the fruits of our efforts to design and implement a sound data collection process, and in Chapter 13 offer recommendations for the conduct of future work.

Chapter 12

Findings on the Administration of the Observational Session

Carolyn A. Eldred

This chapter describes and assesses various aspects of the observational sessions as they were actually administered, based on empirical data from interviewer reports and ratings, maternal self-reports, variables coded in the university laboratories, and comments about the session and its administration recorded by the coders of the videotapes.

Examination of the environment for conducting the sessions revealed that the ideal environment for data collection was rarely encountered: the sessions were typically conducted with persons besides the mother and child present, interruptions and distractions were commonplace, and the work space was not always optimal.

Interviewer performance was examined for the two of the five tasks in the protocol for which systematic coders' comments were available. Based on these two tasks, some deviation from the protocol was observed in 42 percent of the cases. Nonetheless, if the two tasks are viewed as a sample of interviewer performance, it appears that a large majority of the sessions were administered satisfactorily, if not optimally: in 80 percent of the sessions, either a relatively minor deviation was observed on only one of the two tasks or no deviations were noted at all. Deviations from the intended procedures ranged from those attributable to ambiguities in the study protocol (and perhaps to some of its more ambitious requirements) to a small minority that represented a dramatic departure from the principles we had attempted to instill (4 and 6 percent, respectively, for the two tasks examined).

Based on a self-report measure of the mothers' experience of the session, we conclude that their experience was generally positive, with 54 percent rating it in the "very positive" range and virtually all of the rest rating it in the intermediate range. Nonetheless, 15 percent gave quite negative ratings when focusing more narrowly on the challenges presented by the situation. Other measures of the mothers' comfort and functioning in the situation that were coded from the tapes or rated by the interviewers were strongly related both to maternal self-reports and to one another.

The age of the child was the central determinant of the experience of the session for both mother and child: the experience was more negative the younger the child, presumably because the tasks posed greater challenges. In addition, interviewers made more errors when the children were younger.

Having described how the observational work was designed and implemented, this chapter describes and assesses the conduct of the observational sessions based on empirical data.

Of particular interest are the challenges that faced the data collection teams and the performance of survey interviewers in a nontraditional role. As noted in previous chapters, this work required them to perform activities not usually associated with a survey interaction. But more important than the novelty of the assignment *per se* was the need for them to create a standardized, structured situation that would serve as a blank slate on which the mother and child would engrave their own unique response. This implied administering the protocol in such a way that all the mothers began with a clear idea of the objectives of each task and maintaining a polite professionalism and neutral stance while each dyad worked together. The challenges to mothers and children of working together on tasks that required some "stretch" for the children were not to be magnified by problems created by the interviewer, such as failing to explain a task adequately or creating additional anxiety by emphasizing the child's performance. Nor were the inherent challenges of the tasks to be reduced by "helpful" suggestions from the interviewer. Other extraneous influences that we had hoped to minimize were intrusions by other people or events during the observational session. Clearly, these objectives were, and are, sensible. But at the start of the study, it was unclear how well they would be achieved and how much difference it would make if they were not. This chapter attempts to explore these issues.

The following sections of this chapter draw on a variety of data sources: interviewer reports and ratings, maternal self-reports, comments about the session and its administration recorded by the university coders, and variables coded in the university laboratories. Included are descriptors of the environments in which the sessions were conducted, indicators of interviewer performance, and subjective reports of maternal comfort or distress during the session. In addition, several substantive measures from the university laboratories, such as Mother's Confidence, were included, because of their potential sensitivity to the emotional ambience created by the interviewer and her skill at explaining and structuring the tasks.

I. The Environment for Conducting the Session

In this section we present information about the observational sessions conducted for the study, based on reports from the interviewers who conducted them and comments of the university coders who reviewed the videotapes. Interviewers systematically recorded information about the presence of other individuals during the session and the occurrence of a variety of distractions and interruptions. In addition, in the course of making their substantive ratings, the Harvard team's coders wrote short narratives describing what was occurring during the two activities they coded (the book reading activity and wheels task). These narratives form the basis for some of the information about the environment for conducting the sessions that is presented in this section, as well as the assessment of interviewer performance presented in the next section.

As previously noted, the pretests provided a preview of what could be expected when taking observational work out of the laboratory and into the homes of young, disadvantaged women. Just as in the pretests, other adults and children were frequently present during the session, there were a variety of distractions and interruptions, and optimal workspace was not always available.

While interviewers tried to schedule the sessions when only the mother and the focal child were at home, they were generally unable to do so: in 70.9 percent of the cases, there was someone else present. Tables 12.1 and 12.2 present data based on systematic interviewer reports concerning the presence of other people in the immediate vicinity ("within view or earshot") of the observational session. Children other than the focal child were most common, being present in 51.0 percent of the cases. In a little more than half the cases in which another child was present, another adult was in the vicinity; but in the remaining cases, or 23.6 percent of the total number completed, at least one child was present but no adult (other than the mother and data collection team) to distract or attend to the child. Table 12.1 also indicates the presence of a variety of other people who were sometimes present during the session.

Although the presence of others did not guarantee interruptions of the session, this was often the effect. Table 12.3, also based on systematic interviewer reports, presents the data. In 33.8 percent of the sessions another child or children interrupted the session at least once; in 7.6 percent of completed sessions there were three or more such interruptions; 25.2 percent of the sessions were interrupted at least once by an adult, and someone entered or left the home during 39.3 percent of the sessions. Telephone calls involving the mother interrupted 18.6 percent of the sessions. In addition, background noise was often present: in 46.6 percent of the cases, there was audible background conversation, while a radio, television, or stereo contributed background noise in 29.3 percent of the cases. Interviewers in 2.1 percent of the cases volunteered that noise from the street or outside the building was distracting. Focal children themselves could be responsible for interrupting the observational session, for instance, by needing a bathroom break or wandering away from the work area; interviewers volunteered that this occurred in 4.8 percent of the completed cases.

The need to adapt to the realities of the home situation extended to selecting or creating a place conducive to both completion of the activities and videotaping. Interviewers were responsible for working out a suitable arrangement with the mother once the data collection team was in the home, but they were not asked to report on the specific arrangements for each session. However, coders working with the Harvard team routinely recorded comments regarding the conduct of the sessions; in about five-sixths of the completed cases (243 cases) these comments included some mention of seating arrangements and workspace for the book reading and wheels tasks, although the level of detail varied from case to case.

Ideally, the mother and child were to sit side by side for the first activity—book reading and discussion — or the child could sit in the mother's lap if this occurred spontaneously. For the remaining activities — the teaching tasks — the mother and child were to sit at a 90-degree angle to one another, ideally at adjacent sides of a kitchen or dining table, although a coffee table or even the floor could be used if necessary. It would be difficult for the mother to work with her child on these activities if the child were still in her lap, however.

In about half of the cases with commentary mentioning seating arrangements (53.1 percent), the Harvard coders explicitly noted that the dyad was seated at a table, as shown in Table 12.4. In about one-fourth of the cases with commentary (26.7 percent), the dyad was described as sitting on a couch or in chairs; in some of these the use of a coffee table for work space was explicitly noted, but a coffee table may also have been used in some instances in which it was not mentioned. In 14.8 percent of the cases, the dyad worked on the floor, generally for both sets of

Table 12.1

Combinations of Person(s) Other Than Respondent and Focal Child Present During the Observational Session

Person(s) Present	Percent
No one else	29.1
Anyone else	70.9
Child(ren) and other adult(s)	27.4
Child(ren) without other adult(s)	23.6
Other adult(s) only	19.9
Sample size	290

Source: New Chance observational study interviewer reports.

Notes: Calculations for this table used data for all 290 respondents for whom there were 18-month follow-up survey data and a completed observational session, including those with values of zero for outcomes and New Chance enrollees (i.e., experimentals) who did not participate in the program. The sample size may fall slightly short of this number because of missing data.

Distributions may not total 100 percent because of rounding.

Table 12.2

Person(s) Other Than Respondent and Focal Child Present During the Observational Session

Person(s) Present	Percent
Child(ren) other than focal child	51.0
Respondent's husband or boyfriend	13.1
Respondent's mother	14.9
Respondent's father	2.8
Other adult relative(s)	23.2
Friend(s)	10.7
Other or unspecified adult(s)	10.7
Sample size	290

Source: New Chance observational study interviewer reports.

Notes: Calculations for this table used data for all 290 respondents for whom there were 18-month follow-up survey data and a completed observational session, including those with values of zero for outcomes and New Chance enrollees (i.e., experimentals) who did not participate in the program. The sample size may fall slightly short of this number because of missing data.

Entries total more than 100 percent because interviewers recorded the presence or absence of persons in each discrete category shown in the table.

Table 12.3

Interruptions and Distractions During the Observational Session

Interruption or Distraction	Percent
Interruptions by child(ren) other than focal child	33.8
Interruptions by other adult(s)	25.2
Telephone calls involving the respondent	18.6
Someone entering or leaving the home	39.3
Background conversation	46.6
Noise from radio, television, or stereo	29.3
Sample size	290

Source: New Chance observational study interviewer reports.

Notes: Calculations for this table used data for all 290 respondents for whom there were 18-month follow-up survey data and a completed observational session, including those with values of zero for outcomes and New Chance enrollees (i.e., experimentals) who did not participate in the program. The sample size may fall slightly short of this number because of missing data.

Entries total more than 100 percent because interviewers recorded the occurrence or nonoccurrence of interruptions/distractions in each discrete category shown in the table.

Table 12.4

Seating Arrangements of Mother and Child During Observational Session

Arrangement	Percent
At a kitchen or dining table	53.1
On a couch or in chairs	26.7
On the floor	14.8
On a bed	1.2
Location not noted except that child was in mother's lap	3.3
Other	0.8
Sample size	243

Source: Author's content analysis of Harvard coders' comments on administration of the observational session.

Notes: Calculations for this table used data for 243 respondents for whom there were 18-month follow-up survey data, a completed observational session, and commentary describing seating arrangements, including those with values of zero for outcomes and New Chance enrollees (i.e., experimentals) who did not participate in the program.

Distribution does not total 100 percent because of rounding.

tasks; in a few cases, however, the dyad performed the book reading and discussion seated on a couch or in chairs and then moved to the floor for the teaching tasks.

Because the need for a different seating arrangement for the book reading and teaching tasks presented challenges for the data collection team, we wondered how well the team carried out the intent of the absent researchers (see Chapter 10). As the second teaching task, the wheels task provided a clue to the interviewers' success in rearranging the pair following the book reading activity. While information was not available about specific seating arrangements around the work space, the coders did note when the child sat in the mother's lap. As mentioned above, this was considered an acceptable arrangement for the book reading activity but not at all for the teaching tasks. It is noteworthy that lap sitting did diminish, but not disappear entirely, between the book reading activity and the teaching tasks. In 18.5 percent of the cases with commentary about seating arrangements, the child was described as sitting on the mother's lap for the book reading and discussion; this frequency dropped to 7.4 percent for the wheels task. Interestingly, the Harvard coders noted a plausible explanation for this arrangement during the wheels task in about a third of the cases in which it occurred (for example, sitting in the mother's lap helped the child to focus on the task after becoming distracted). In such cases, the interviewer may have made a conscious judgment that the advantages of lap sitting outweighed the disadvantages and decided not to intervene.

Anecdotal material in the comments of the Harvard coders was a source of information about the realities of conducting this work in the home environment, including occasional unintended consequences of some procedural decisions made by the observational study team. For instance, the procedures provided for the videographer to interact with the child while the interviewer explained the tasks to the mother, including entertaining the child with a "Sesame Street" hand puppet. In a few cases the child continued to display an active interest in the videographer and/or the puppet as the mother was trying to interest him or her in the tasks; in one instance, the child held the puppet throughout the session. (Whether or not the interviewer or videographer attempted to do anything about this was unclear.) During the wheels task, children sometimes left the work area to look for toys or other objects with wheels, sometimes with the mother's encouragement. Finally, other adults or children could intervene in the activities, sometimes making "helpful" suggestions and other times making fun of what the mother or child was saying or doing. In a few instances another child or a baby was in the immediate work area but was not an active participant.

In summary, as foreshadowed by the pretests, the ideal environment for data collection was rarely encountered: the sessions were typically conducted with persons besides the mother and child present, interruptions and distractions were commonplace, and the work space was not always optimal. Thus, there were limits on the data collection teams' ability to structure the setting, and they had to be prepared to be flexible — and sometimes creative — in shaping and managing the physical and social environment.

II. Interviewer Performance

A. Indicators of Interviewer Performance

The commentary of the Harvard coders was also a source of information about how well the observational session was conducted and what kinds of problems occurred. While their comments primarily addressed the mother-child interaction, problems in administration of the session were also regularly noted. Comments pertinent to administration of both the book reading and discussion activity and the wheels task were abstracted from the commentaries for each case and subjected to a content analysis to categorize the coders' criticisms about administration of these tasks. The problems were characterized by both type and severity.

It is important to understand that the coders' comments reflected their reactions to what they observed on the tapes rather than a systematic comparison of what they saw with the instructions in the script (questionnaire) and procedures manual. While most of the problems cited did involve errors, or at least poor judgment, on the part of interviewers, some of the behaviors the coders found inappropriate appeared to result from interviewers' inflexibility in rigidly *adhering* to the script and procedures. Thus, the raters' comments elucidate not only interviewer performance but also the survey procedures themselves, viewed against the expectations of laboratory-based researchers for survey-based data collection.

However, a limitation of the raters' comments is that they do not derive from a systematic checklist, for example, but from narrative comments relating to several general issues of protocol administration.[1] Because of different thresholds and styles of recording and the fact that the typology to categorize the comments was imposed after the fact (rather than being available to the coders), there may be some unevenness in the reporting. In particular, the absence of a particular mention carries a somewhat equivocal meaning.[2]

Another potential limitation is that the focus of this review was on negative aspects of performance; since the coders' comments were not intended as a mechanism for assessing interviewer performance, positive behaviors were generally not noted. Because we had no way of identifying exemplary performance, if it existed, the discussion below may underestimate the variability of interviewer performance. Furthermore, we do not have comparative data for academically trained data collection personnel, such as graduate research assistants working in a child development laboratory, for whom administering protocols like that used in this study may come more "naturally" than it does to a survey interviewer. Yet even academically trained personnel could be expected to make some of the same kinds of errors observed among survey interviewers as they try to master an unfamiliar protocol.

In addition, while it was hoped that review of this narrative commentary would shed some light on the ability of survey interviewers to maintain the integrity of an observational protocol outside the laboratory, available resources dictated that the content analysis be carried out by one team member only, without an opportunity to establish interrater reliabilities. On the positive side, this team member, who had watched many tapes and was therefore familiar with the performance of individual interviewers, was blind with respect to interviewer identity. During the content analysis, she was also blind to the various measures of the mother's subjective experience of the session that would subsequently be examined in assessing the importance of interviewer performance.[3]

---

[1] This is a limitation shared by open-ended questions in survey questionnaires: specific items of information are "volunteered" by a respondent and a typology imposed after the fact. But, in the absence of a specific prompt, thresholds for giving a particular answer may vary even among those to whom it applies equally.

[2] The team had not originally planned on such a use for these comments, and it was not feasible to review the tapes again, after they had already been coded, simply to assess interviewer performance.

B. Deviations from the Observational Session Protocol

Table 12.5 categorizes the types of deviations described by the Harvard coders for each of the two tasks coded by the Harvard team — book reading and wheels. Overall, deviations were noted for 42.0 percent of the cases on either the book reading or wheels task; in 30.1 percent there were problems connected with the book reading task, and in 21.7 percent with the wheels task. The difference in the frequency of deviations on the two tasks is in large part a function of a problem that was applicable only to the book reading task — inappropriate handling of the reminder to discuss the book, which occurred in 8.9 percent of the cases.

Many of the deviations cited by the coders represent errors by the interviewers, in that they disregarded instructions or principles contained in the various study materials and addressed in interviewer training. The most common of these (in 8.9 percent of the book reading cases and 7.2 percent of the wheels cases) involved offering direction, intervening in the dyad's work, or otherwise abandoning a neutral stance in some way. Other outright errors, occurring with considerably less frequency, included being either negative (for example, condescending or annoyed) or overly familiar with the mother or child, allowing overlap between tasks (for example, giving instructions or assembling materials for one task while another was under way), or failing to tape the initial instructions to the mother.[4]

Other problems involved poor delivery, that is, a presentation that was unnatural or stilted, was misleading or confusing, or failed in other ways to provide clear, accurate information.

A final group of problems, which occurred with considerable frequency, seemed to reflect difficulties with the "interface" between data collection protocol and the interviewer. One type of problem involved the discussion that was to take place at the conclusion of the book reading task; this problem was noted in 8.9 percent of the cases. The instructions for this task asked the mother to "read or look at the book" with her child and then "take a few minutes to talk about the book" with the child afterwards. If the mother did not spontaneously talk with her child after going through the book, the interviewer was to remind her to do so.[5] While the coders' comments suggested that interviewers were sometimes simply careless or negligent in failing to prompt the mother to discuss the book, in most cases it appeared that the interviewers were having trouble

---

[3] The three Harvard variables related to the mother's experience of the session and the child's age were included with the coders' comments, but were not extracted or available during the content analysis.

[4] The last presents a problem primarily for those who view the tapes because they lack the context provided by the explanation to the mother. They are unable to evaluate its adequacy or assess the mother's understanding and reaction to the task. However, failure to tape the initial instructions also deprives the mother of an opportunity to acclimate to being taped before beginning to work with her child.

[5] An earlier version of the instructions, used in the first several sessions completed, delivered a reminder to discuss the book only if the mother was still going through the book toward the end of the allotted time and there was a chance she would not have time for discussion. These earlier instructions required the interviewer to interrupt the mother and ask her to stop and discuss the book; while some of the coders' comments on the early cases seemed to reflect the awkwardness of this particular procedure, the discussion instruction presented an ongoing challenge.

Table 12.5

Types of Deviations in Session Administration Noted in Coders' Comments, for the Book Reading and Wheels Tasks

Type of Deviation	Book Reading (%)	Wheels (%)
Any deviation mentioned	30.1	21.7
Unnatural presentation or interaction	4.5	1.4
Misleading or confusing presentation or interaction	2.4	2.1
Directive or intervening behavior	8.9	7.2
Negative demeanor (e.g., curt or condescending)	1.4	0.7
Overly familiar demeanor	2.1	2.8
Poor management of timing and flow of tasks	7.9	9.0
Failure to make clean transitions between tasks	0.3	2.1
Inappropriate handling of reminder to discuss the book	8.9	N/A
Failure to tape initial instructions to the mother	2.1	1.7
Other, unclassified, deviation	2.1	0.0
Sample size		292

Source: Author's content analysis of Harvard coders' comments on administration of the observational session.

Notes: Calculations for this table used data for 292 respondents for whom there were 18-month follow-up survey data, a completed observational session, and commentary about task administration (including two with usable audiotapes but unusable videotapes who are therefore omitted from most other tables in the monograph), including those with values of zero for outcomes and New Chance enrollees (i.e., experimentals) who did not participate in the program.

Sum of subtypes exceeds frequency of "any" deviations, because multiple deviations per task per case were sometimes observed.

determining what constituted a discussion and whether or not it had occurred. Thus, the coders' criticisms took two forms: (1) interviewers were prompting discussion when it had already occurred or (2) they were failing to prompt discussion when the mother had not discussed the book spontaneously. There was considerable variation in the mothers' approaches to the book reading activity itself; some engaged in considerable discussion while going through the book or segued "seamlessly" into discussion while finishing the story, thus creating some ambiguity as to whether or not a discussion *per se* had already taken place. In other cases the mother simply asked if the child liked the book, and the child gave a one-word reply, but whether this interchange should "count" as a discussion was unclear even in the coders' comments.

A related set of problems was labeled "poor time management" and generally included situations in which the coders perceived something inappropriate about the way the interviewer managed the timing of the tasks. For instance, the interviewer may have terminated the task very abruptly or just as the dyad was at a critical point, or she may have let it drag on for too long, creating an awkward pause or seeming to be insensitive to the wishes of the mother and child to move on. Such problems were noted in 7.9 percent of the cases for the book reading task and in 9.0 percent for the wheels task. Clearly, some of these problems resulted from failure to follow the time guidelines specified in the script; however, others probably resulted from the interviewer's attempt to follow them to the letter.[6]

As shown in Table 12.6, an attempt was also made to characterize the severity of problems noted for the book reading and wheels tasks. A single score was assigned to each task for each case based on the overall commentary of the coders (regardless of the number or type of deviations noted for the task). Of the sessions with problems on each task, 20.2 percent of the book reading problems and 12.2 percent of the wheels problems were classified as "noteworthy but not serious" (the instructions were clear but awkwardly delivered, for example). Such problems were observed in 20.2 percent of the completed book reading tasks and in 12.1 percent of the wheels tasks overall. "Moderately serious" deviations included inappropriate responses — generally of a reactive nature — such as giving a nonneutral answer to a mother's request for clarification; such problems occurred in 5.5 percent of the book reading tasks and in 4.1 percent of the wheels tasks. "Very serious" deviations included instances of active or gratuitous intervention or behaviors that could seriously undermine the dyad's composure, motivation, or understanding and performance of a task (for example, prompting the book discussion by telling the mother to "ask him some questions"). Deviations of this severity were cited in 4.1 percent of the book reading tasks and in 4.5 percent of the wheels tasks. Finally, as noted in the table, there were just a few instances of "extremely serious" departures from protocol or patently offensive behavior (for example, with the videotape turned off and the audiotape running, the interviewer, videographer, and mother discussed strategies to help the child name wheeled objects).

In addition to the severity score for each of the two tasks, ranging from 0 to 4 (0 indicated

---

[6] Timing the tasks required the interviewer to record the start time and a target completion time for each task. However, she was not provided a watch with a sweep-second hand for this purpose, so timing could not be very precise. One task (wheels) could be concluded early if the child did not correctly name any objects during the first 2.5 minutes. Interviewers were also permitted to go on if the child completed a task and there was nothing left to be done (for example, exhausted all of the available blocks in replicating the model). Otherwise, interviewers were to allow each task to run for its allotted time.

Table 12.6

Severity of Deviations in Session Administration Noted in Coders' Comments, for the Book Reading and Wheels Tasks

Level of Severity[a]	Book Reading (%)	Wheels (%)
Mean	0.5	0.4
Standard deviation	0.8	0.8
0 (no deviation mentioned)	69.9	78.3
1 (noteworthy, but not serious)	20.2	12.1
2 (moderately serious)	5.5	4.1
3 (very serious)	4.1	4.5
4 (extremely serious)	0.3	1.0
Sample size		292

Source: Author's content analysis of Harvard coders' comments on administration of the observational session.

Notes: Calculations for this table used data for 292 respondents for whom there were 18-month follow-up survey data, a completed observational session, and commentary about task administration (including two with usable audiotapes but unusable videotapes who are therefore omitted from most other tables in the monograph), including those with values of zero for outcomes and New Chance enrollees (i.e.,experimentals) who did not participate in the program.

Distributions may not total 100.0 percent because of rounding.

[a] Severity scores reflect the seriousness of the overall picture of observed deviations for each task, whether single or multiple. Examples of each severity level: (1) noteworthy but not serious--instructions clear but awkwardly delivered; (2) moderately serious--reacting inappropriately to the mother's query or behavior, for example, by agreeing with the mother's suggestion about what she should do rather than remaining noncommittal; (3) very serious--actively and gratuitously intervening in the interaction or engaging in any behavior that could seriously undermine the dyad's composure, motivation, or understanding and performance of the task; (4) extremely serious--an extreme example of a "very serious" deviation above (e.g., with the tape turned off), mother, interviewer, and videographer join in a discussion about how to get the child to name wheeled objects.

BEST COPY AVAILABLE

that no problem was cited), an overall Severity of Deviations score was created by summing the scores on the two tasks. This overall score could theoretically range from 0 to 8, but in actuality did not exceed 6, with a mean of .828 and a standard deviation of 1.28 across the completed cases. In addition to the caveats noted earlier, this summary score is somewhat limited by the fact that it is based on only two of the five activities that constituted the session — the book reading and discussion, which introduced the session, and the wheels task, the second of the four teaching tasks. For this and previously mentioned reasons, a score of 0 cannot be interpreted as indicating the absence of problems. Still, if these two tasks are viewed as samples of the interviewer's behavior during the session, the summary Severity of Deviations score becomes a proxy for overall interviewer performance. As such, it allows us to begin to assess how well interviewers mastered the task of administering the observational protocol.

C. Variation in Interviewer Performance

Table 12.7 indicates the proportion of completed cases associated with each overall Severity of Deviations score. As noted in the table, 58.0 percent of the completed cases had a summary score of 0, meaning that the coders noted no deviations on either of the tasks. Another 21.9 percent of the cases (or a little more than half in which any deviation was noted) had a score of 1, meaning that the coders noted a "noteworthy, but not serious" deviation on one, but not both, of the tasks. Together, cases in these two categories accounted for four-fifths of the completed sessions. Still, there is room for concern about the remaining cases, which were characterized by more, or more serious, deviations.

The contribution of individual interviewers to the overall pattern of performance is of interest, particularly since interviewers varied greatly with respect to the number of cases they completed, ranging from 1 to 52. Not surprisingly, their typical performance also varied, as inferred from the average of the summary Severity of Deviations scores for each interviewer, which ranged from 0 to 3.25. Table 12.8 shows the percentage of interviewers with average scores at each level and the percentage of completed cases for which these interviewers were responsible. Table 12.9 categorizes the interviewers according to the number of cases they completed; it also presents the range of the average severity scores of the interviewers in each size group as well as the overall average of the interviewers in each group. Together, these two tables indicate a positive association between interviewer proficiency (as inferred from ratings based on the coders' comments for the book reading and wheels tasks) and volume of cases. That is, more proficient interviewers completed more cases. Whether this is a function of retraining and practice effects or the elimination of the less proficient (or less committed) interviewers, the news is encouraging.

Given the variation in interviewer performance, a natural question is whether any particular interviewer attributes were associated with the quality of the work. As previously noted, few survey efforts are staffed in a way that will support the systematic investigation of interviewer effects, and the observational study in particular was characterized by a distribution of interviewer assignments that dictated caution in attempting to investigate interviewer effects. Nonetheless, an exploratory regression analysis was carried out that examined Interviewer's Age, Interviewer's Education, Ethnic Match of Interviewer and Respondent, and Interviewer's New Chance Experience as possible predictors of the Severity of Deviations score. The findings did not suggest any association between these variables and interviewer proficiency.

Table 12.7

Distribution of Completed Observational Sessions, by Summary Severity of Deviations Scores for the Book Reading and Wheels Tasks

Summary Severity of Deviations Scores[a]	Percent
0 (no deviations noted)	58.0
1	21.9
2	8.0
3	6.9
4	2.4
5	2.1
6 (deviations on both tasks, at least one of which was very or extremely serious)	0.6
Sample size	288

Source: Author's content analysis of Harvard coders' comments on administration of the observational session.

Notes: Calculations for this table used data for 288 respondents for whom there were 18-month follow-up survey data, a completed observational session, and a commentary file for both the book reading and wheels tasks, including those with values of zero for outcomes and New Chance enrollees (i.e., experimentals) who did not participate in the program.

Distribution does not total 100 percent because of rounding.

[a] Calculated by summing the scores on the book reading and wheels tasks. Summary scores could theoretically range from 0 to 8 but actually ranged from 0 to 6.

Table 12.8

Distribution of Interviewers and Observational Sessions, by Interviewer's Average Summary Severity of Deviations Score

Interviewer's Average Summary Severity of Deviations Score[a]	Interviewers (%)	Sessions Completed by Interviewers Whose Average Score is in This Range (%)
< .50	26.9	24.8
.50 - .75	19.2	23.1
.76 - .99	11.5	15.2
1.00 - 1.25	19.2	27.9
1.26 - 1.99	7.7	4.1
2.00 or higher	15.4	4.8
Sample size	26	290

Sources: Author's content analysis of Harvard coders' comments on administration of the observational session and interviewer information in New Chance observational study data file.

Notes: Calculations for this table used data for all 290 respondents for whom there were 18-month follow-up survey data, a completed observational session, and commentary on task administration, including those with values of zero for outcomes and New Chance enrollees (i.e., experimentals) who did not participate in the program.

Distributions may not total 100 percent because of rounding.

[a] Calculated by summing the scores on the book reading and wheels tasks.

Table 12.9

Average Summary Severity of Deviations Scores of Interviewers, by Number of Sessions Completed

Number of Sessions Completed	Interviewers (%)	Each Interviewer's Average Summary Severity of Deviations Score[a]	
		Range	Average
20 or more	15.4	.23 - 1.06	0.67
10-19	23.1	.20 - 1.18	0.74
5-9	38.5	0 - 2.40	0.84
4 or fewer	23.1	0 - 3.25	1.63
Sample size			290

Sources: Author's content analysis of Harvard coders' comments on administration of the observational session and interviewer information in New Chance observational study data file.

Notes: Calculations for this table used data for all 290 respondents for whom there were 18-month follow-up survey data, a completed observational session, and commentary on task administration, including those with values of zero for outcomes and New Chance enrollees (i.e., experimentals) who did not participate in the program.

Distribution does not total 100 percent because of rounding.

[a] Calculated by summing the scores on the book reading and wheels tasks.

Interestingly, even interviewers who had predominantly low Severity of Deviations scores for their cases often had an occasional case with a high score. This might suggest that even those who are generally proficient may react less than optimally when confronted with greater situational challenges. One of these may be the age of the child, which, as discussed later, is associated with several measures of the participants' experience of the session. Because the teaching tasks were designed to be moderately challenging, mothers of younger children are likely to face greater challenges in working with their children, other things being equal. And such challenges might translate into challenges for the interviewer, who may succumb to inappropriate behavior in handling more difficult situations. Indeed, a small but significant association in the predicted direction was observed between Child's Age and Severity of Deviations ($r = -.168$, $p < .01$). That is, cases involving younger children were characterized by more severe deviations from the protocol.

D. An Overall Assessment of Interviewer Performance

Comments from the Harvard laboratory on the two tasks coded there — the book reading and discussion activity and the wheels task — indicate that there was room for improvement in the interviewers' performance: in 42.0 percent of the sessions, the coders noted some deviation from the protocol. However, in about half of these cases, only a relatively minor deviation on one of the two tasks was noted. Still, a deviation judged to be at least moderately serious was noted in about 10 percent of administrations of each task. Some interviewer errors seemed to result from disregarding the study procedures and/or from administering sessions without sufficient review of the procedures. Other problems arose either because interviewers were expected to make difficult judgments in the press of the situation or because they were discouraged from exercising their own judgment, sometimes resulting in behavior that appeared awkward or inappropriate when viewed on tape. Nonetheless, if the coders' comments for these two tasks can be viewed as indicative of the overall proficiency with which the protocol was administered, the vast majority of sessions seem to have been administered satisfactorily even if not optimally: 79.9 percent were characterized by either no deviations or only a relatively minor deviation on one task.

III. Subjects' Experience of the Session

A. An Overview of the Measures

Having considered at length how the session was administered, we now consider how it was experienced by the study subjects. Data bearing on this issue come from a variety of sources. First, the mothers spoke for themselves in their responses to a self-administered questionnaire they completed at the close of the session. In addition, the interviewers made ratings of how self-conscious the mother and child appeared during the session. Finally, a few of the ratings made by each university team were selected for their methodological interest, because they were potentially susceptible to the influence of the interviewer and the immediate environment. They were included because review of the videotapes and the comments of the Harvard coders pointed to the role of the interviewers in setting the affective tone of the session, providing the mother the cognitive tools she needed to work with her child effectively on the tasks, and, through these mechanisms, possibly contributing to the way in which mother and child behaved and felt during the session. These measures included the Harvard team's Mother's Comfort Level While Reading, Mother's Comfort Level During Wheels Task, and Mother's Ease of Ideas to Keep Wheels Task Going, as well as the Minnesota team's Mother's Confidence and Child's Experience of the Session. The ratings from each of the four sources — mothers, interviewers, and the two university laboratories — were made independently of one another.

Following discussion of these measures in this section, the chapter concludes with an examination of the various factors that may have contributed to subjects' experience of the session, including the performance of the interviewers.

B. Mothers' and Children's Subjective Experience of the Session

Because of sensitivity to the potential intrusiveness of a study like this, as well as concern about possible reactivity associated with taping the session, we were particularly interested in the self-reported reactions of the mothers. Respondents' experience of the session was measured in a 10-item self-administered questionnaire that asked about such matters as the mother's anxiety about her own or her child's performance, her enjoyment of the tasks, and the interviewer's role in helping her feel comfortable.[7] (These items are listed in Table 12.10.)[8] Of particular interest was whether the experience was especially unpleasant or stressful, so the items were designed to tap those aspects of the situation suspected to be sources of stress.

Distributions and means for each of the maternal ratings about the session appear in Table 12.10. The table indicates considerable variation among the items in their discriminatory power, with the statements in which disagreement indicated well-being performing better on the whole. Interviewers received rather high marks for helping the mothers to feel comfortable, and most mothers said that they had enjoyed working with their children and that they were reasonably relaxed during the session. At the same time, they found it somewhat difficult to forget about the videotaping completely, and there was some anxiety about performance during the session. Interestingly, although the interviewers had found distractions and interruptions of the session to be prevalent (as discussed earlier in this chapter), the mothers themselves were not particularly bothered by them.

The 10 items in Table 12.10 were summed to form a scale, after recoding or "reflecting" the ratings for the items in which agreement indicated a low level of well-being so that a higher score would uniformly indicate a greater subjective well-being. Overall, the 10-item scale yielded a moderate standardized Cronbach's alpha of 0.69.[9] Two subscales, representing the positive aspects of the experience (the first five statements in Table 12.10) and the negative aspects (the second five statements), had alphas of .65 and .69, respectively.

---

[7] An interviewer-administered open-ended question provided an additional opportunity to reflect on the session, but without the privacy afforded by the self-administered questionnaire. The majority of respondents did not comment at all in response to this question, and the answers of those who did conveyed little additional information about respondents' perceptions of the session.

[8] In the table the "positive" statements have been grouped together and are followed by the "negative" statements. The original order is indicated by the letter preceding each statement.

[9] In calculating overall scale scores, cases missing three or more items were dropped, resulting in N = 287. In the item analysis, cases missing any items were dropped, resulting in N = 270.

Table 12.10

Maternal Self-Reports of the Subjective Experience of the Observational Session

Statement Rated by Respondent[a]	Average or Percent
Statements for Which Agreement Indicates Positive or Pleasant Experience	
a. I forgot all about the videotaping once I started working with my child (average rating)	5.91
0 - 2 (low agreement) (%)	16.4
3 - 7 (moderate agreement) (%)	47.4
8 - 10 (high agreement) (%)	36.2
c. The interviewer helped me feel comfortable during the session (average rating)	8.42
0 - 2 (low agreement) (%)	3.1
3 - 7 (moderate agreement) (%)	20.4
8 - 10 (high agreement) (%)	76.5
e. I enjoyed reading or looking at the book with my child (average rating)	9.11
0 - 2 (low agreement) (%)	1.7
3 - 7 (moderate agreement) (%)	10.2
8 - 10 (high agreement) (%)	88.1
h. I enjoyed showing my child the games (average rating)	9.21
0 - 2 (low agreement) (%)	1.8
3 - 7 (moderate agreement) (%)	10.7
8 - 10 (high agreement) (%)	87.5
j. I felt relaxed during the session (average rating)	7.61
0 - 2 (low agreement) (%)	5.2
3 - 7 (moderate agreement) (%)	33.3
8 - 10 (high agreement) (%)	61.5
Statements for Which Agreement Indicates Negative or Stressful Experience	
b. It was hard for me to figure out how to get my child to do the games (average rating)	3.66
0 - 2 (low agreement) (%)	47.9
3 - 7 (moderate agreement) (%)	33.3
8 - 10 (high agreement) (%)	18.8
d. While I was working with my child on the book and games, I felt nervous about how I was doing (average rating)	4.40
0 - 2 (low agreement) (%)	41.9
3 - 7 (moderate agreement) (%)	30.6
8 - 10 (high agreement) (%)	27.5

(continued)

Table 12.10 (continued)

Statement Rated by Respondent[a]	Average or Percent
f. Other things going on in my home made it difficult to listen to the interviewer's instructions or work with my child (average rating)	2.95
0 - 2 (low agreement) (%)	57.8
3 - 7 (moderate agreement) (%)	24.4
8 - 10 (high agreement) (%)	17.8
g. I felt nervous about how my child was doing on the games (average rating)	4.63
0 - 2 (low agreement) (%)	38.8
3 - 7 (moderate agreement) (%)	30.4
8 - 10 (high agreement) (%)	30.8
i. The games were difficult for my child (average rating)	3.94
0 - 2 (low agreement) (%)	39.0
3 - 7 (moderate agreement) (%)	45.3
8 - 10 (high agreement) (%)	15.7
Sample size	290

Source: New Chance brief interview accompanying observational session.

Notes: Calculations for this table used data for all 290 respondents for whom there were 18-month follow-up survey data and a completed observational session, including those with values of zero for outcomes and New Chance enrollees (i.e., experimentals) who did not participate in the program. The sample size may fall slightly short of this number because of missing data.

In this table, the positive statements have been grouped together and are followed by the "negative" statements. The original order is indicated by the letter preceding each statement.

Distributions may not total 100.0 percent because of rounding.

[a] Maternal ratings can range from 0 to 10, in which 0 means "not at all true" and 10 means "completely true."

Table 12.11 shows the distribution of completed cases on the maternal self-report scales, along with the measures related to the experience of the session based on the remaining sources: Harvard raters, Minnesota raters, and interviewers. Based on the 10-item scale score, Mother's Overall Subjective Well-Being During Session, the session was generally a positive experience: 54.1 percent rated it in the "very positive" range, and virtually all the rest rated it in the "intermediate" range. A slightly different picture emerges when the two subscales are considered separately, however. When the potentially positive aspects of the experience are viewed separately, 76.0 percent rated the session as "very positive." But when the potentially stressful aspects are viewed separately (for example, nervousness, difficulty working with the child), only 37.9 percent reported a very positive experience, and 14.6 percent rated the experience as "very negative." The items assessing the stress or challenge of the situation tended to focus more directly on performance of the tasks themselves than the items assessing positive aspects of the experience, which assessed affect during the session more globally. So despite hearing from some mothers that they found the tasks somewhat difficult or anxiety-producing, it is encouraging to find that those feelings did not preclude having a generally positive experience of the session.

Other measures of the mother's subjective experience came from the observations of her behavior during the session, as coded by the Harvard team. This examination of maternal comfort level parallels to a great extent the distribution of the mothers' self-reports on the Negative Experience subscale. Mother's Comfort Level While Reading was rated as "high" for 37.2 percent of the mothers, and Mother's Comfort Level During Wheels Task was rated as "high" for 21.7 percent. Comfort levels for the two tasks were described as low in 10.0 percent and 16.2 percent of the cases, respectively. Not surprisingly, ratings of maternal comfort were lower for the wheels task than for the book reading activities. The wheels task had posed considerable challenges to many mothers, especially those with younger children.

C. Other Experiential Aspects of the Session

Feelings of self-consciousness were another aspect of the experience that might be expected, given the presence of the video camera and the novelty of the situation. Self-consciousness was also a subjective experience we had hoped to minimize. The interviewer ratings of Mother's Self-Consciousness and Child's Self-Consciousness shown in Table 12.11 indicate that the dyads were not observed to be particularly self-conscious in the situation. The interviewers placed 51.0 percent of the mothers and 59.7 percent of the children toward the end of the 11-point scale anchored by the label "completely natural." Just 6.2 percent of the mothers and 4.5 percent of the children were placed at the end of the scale anchored by "extremely self-conscious."

Other measures speak to the mother's mastery — and feelings of mastery — in this contrived and challenging situation, along with her ability to make it a positive experience for her child. These experiences and feelings, in turn, may have been determined not only by the resources and history that the dyads brought to the situation but also by the way in which it was structured and presented to them. Important in their own right, such feelings may be viewed as part of a "package" of subjective experiences associated with the observational session. The Minnesota group assessed Mother's Confidence and characterized only 29.3 percent of the mothers as "mostly confident." The remainder were either "somewhat unconfident" (43.5 percent) or "mostly unconfident" (27.2 percent). Child's Experience of the Session, another variable rated by the Minnesota team based on the videotaped interaction during the entire session, reflects the de-

Table 12.11

Indicators of Mothers' and Children's Subjective Experience of the Observational Session

Measure of Well-Being During Session	Average or Percent
Scales Based on Maternal Self-Report	
Mother's Overall Subjective Well-Being During Session (average rating)[a]	7.09
Less than 3 (very negative experience) (%)	0.7
3-7 (intermediate experience) (%)	45.2
More than 7 (very positive experience) (%)	54.1
Mother's Subjective Well-Being: Positive Experience Subscale (average rating)[a]	8.06
Less than 3 (very negative experience) (%)	0.7
3-7 (intermediate experience) (%)	23.3
More than 7 (very positive experience) (%)	76.0
Mother's Subjective Well-Being: Negative Experience Subscale (average rating)[a]	6.10
Less than 3 (very negative experience) (%)	14.6
3-7 (intermediate experience) (%)	47.5
More than 7 (very positive experience) (%)	37.9
Interviewer Ratings	
Mother's Self-Consciousness (average rating)[b]	7.03
0 - 2 (high self-consciousness) (%)	6.2
3 - 7 (moderate self-consciousness) (%)	42.8
8 - 10 (low self-consciousness) (%)	51.0
Child's Self-Consciousness (average rating)[b]	7.28
0 - 2 (high self-consciousness) (%)	4.5
3 - 7 (moderate self-consciousness) (%)	35.8
8 - 10 (low self-consciousness) (%)	59.7
Coded Variables: Harvard Laboratory	
Mother's Comfort Level While Reading (average rating)[c]	2.27
1 Low (%)	10.0
2 Moderate (%)	52.8
3 High (%)	37.2
Mother's Comfort Level During Wheels Task (average rating)	2.06
1 Low (%)	16.2
2 Moderate (%)	62.1
3 High (%)	21.7
Mother's Ease of Ideas to Keep Wheels Task Going (average rating)[d]	2.50
1 Very low (%)	17.6
2 Low (%)	26.6
3 Moderate (%)	43.8
4 High (%)	12.1

(continued)

Table 12.11 (continued)

Measure of Well-Being During Session	Average or Percent
Coded Variables: Minnesota Laboratory	
Mother's Confidence (average rating)[e]	2.02
1 Mostly unconfident (%)	27.2
2 Somewhat unconfident (%)	43.5
3 Mostly confident (%)	29.3
Child's Experience of the Session (average rating)[f]	4.27
1 Very low (%)	3.5
2 Low (%)	6.4
3 Moderately low (%)	12.7
4 Moderate (%)	29.0
5 Moderately high (%)	35.3
6 High (%)	11.3
7 Very high (%)	1.8
Sample size	290

Sources: New Chance coded observational study variables, variables from brief interview accompanying observational session, and observational study interviewer ratings.

Notes: Calculations for this table used data for all 290 respondents for whom there were 18-month follow-up survey data and a completed observational session, including those with values of zero for outcomes and New Chance enrollees (i.e., experimentals) who did not participate in the program. The sample sizes may fall slightly short of this number because of missing data.

Distributions may not total 100 percent because of rounding.

[a] Scores on the overall scale and the two subscales can range from 0 to 10. Negative items have been recoded so that a high score is uniformly positive.

[b] Self-consciousness was rated by interviewers on an 11-point scale ranging from 0 ("extremely self-conscious") to 10 ("completely natural").

[c] Mother's Comfort Level: (1) low--mother appears very uncomfortable; (2) moderate--mother shows few signs of discomfort; (3) high--mother seems relaxed and seems to enjoy the activity.

[d] Mother's Ease of Ideas: (1) very low--mother misunderstands task or tells child what has wheels; (2) low--mother has great difficulty with task; (3) moderate--mother is challenged but successfully performs task; (4) high--mother has many ideas and uses multiple strategies, tailored to the child's responses.

[e] Mother's Confidence: degree to which the mother seems to believe that she can work successfully with the child in the situation and that the child will behave appropriately.

[f] Child's Experience of the Session: degree to which the child's experience in the session probably resulted in feelings of success and competence on the tasks and confidence in having a good relationship with the mother.

gree to which the child experienced the session as an affirmation of both his or her competence and a positive relationship with the mother, despite the challenges of the tasks. For most of the children, the session was a reasonably positive experience, with 48.4 percent having an experience labeled "moderately high," "high," or "very high." At the same time, 22.6 percent of the children had an experience of the session judged to be "moderately low," "low," or "very low."

The Harvard raters assessed Mother's Ease of Ideas to Keep Wheels Task Going and documented the difficulties that this task could pose, with nearly half of the mothers rated "low" or "very low" on Ease of Ideas. Included in this group were 17.6 percent of the mothers who appeared not to understand the task or did not perform it correctly for some other reason. The task may have pushed the limits of the absent researchers' ability to exert influence over administration of the session. To begin with, the task was inherently challenging for the dyads because it required abstract thinking and creativity on the part of the mother, along with the ability to keep the child engaged in the absence of any play materials; these obstacles could assume greater importance with younger children. Given these inherent challenges, it was important for the mothers to understand what they were being asked to do and to feel comfortable enough to put their best effort into it. But interviewers were also challenged — by the need to provide a clear description of the guessing game without suggesting specific strategies to the mothers. Further, they were expected to do so without the props that allowed them to demonstrate the principles behind other tasks *and* without any built-in mechanism for confirming that the mother understood this task (unlike the other teaching tasks).

D. Interrelationships Among the Subjective Experience Measures

One might expect some overlap of the various measures of the subjective experience of the observational session, either because the measures reflect some common underlying dimensions or because they are based on the same behavioral cues. Table 12.12 shows the correlations among the various measures and indicates that there are fairly strong interrelationships both within and across sources. The mothers' self-reports bore a strong relationship to the university coders' assessments, with the strongest "cross-source" correlation being that between Mother's Overall Subjective Well-Being During Session and the Minnesota team's assessment of Child's Experience of the Session ($r = .362$, $p < .01$). It should not be surprising that mothers have positive feelings about their experience of the session when it is also appears to be gratifying for their children, as assessed by independent raters.

The interrelationship of the measures listed in Table 12.11 was examined by subjecting the eight ratings from these sources to a principal components factor analysis. Measures used in the analysis included two from the Minnesota laboratory (Mother's Confidence and Child's Experience of the Session), three from the Harvard laboratory (Mother's Comfort Level While Reading, Mother's Comfort Level During Wheels Task, and Mother's Ease of Ideas to Keep Wheels Task Going). Also included were the two measures based on interviewer ratings (Mother's Self-Consciousness and Child's Self-Consciousness), and the maternal self-report measure Mother's Overall Subjective Well-Being During Session (based on the full 10-item scale). The analysis, after varimax rotation, yielded two factors with eigenvalues greater than one, which together accounted for 48 percent of the total variance. All five laboratory ratings and the mothers' ratings loaded on the first factor (accounting for 31 percent of the variance), while the two interviewer ratings loaded on the second factor (17 percent of the variance). Although the

Table 12.12

Correlations Among Subjective Experience Measures

Measure	Correlation[a]
Intercorrelations of Harvard Variables	
Mother's Comfort Level While Reading and Mother's Comfort Level During Wheels Task	0.424 ***
Mother's Comfort Level During Wheels Task and Ease of Ideas to Keep Wheels Task Going	0.416 ***
Mother's Comfort Level During Reading and Mother's Ease of Ideas to Keep Wheels Task Going	0.192 ***
Intercorrelation of Minnesota Variables	
Mother's Confidence and Child's Experience of the Session	0.385 ***
Intercorrelation of Interviewer Ratings	
Mother's Self-Consciousness and Child's Self-Consciousness	0.325 ***
Intercorrelations of Harvard and Minnesota Variables	
Mother's Confidence and Mother's	
Comfort Level While Reading	0.317 ***
Comfort Level During Wheels Task	0.341 ***
Ease of Ideas to Keep Wheels Task Going	0.223 ***
Child's Experience of the Session and Mother's	
Comfort Level While Reading	0.296 ***
Comfort Level During Wheels Task	0.311 ***
Ease of Ideas to Keep Wheels Task Going	0.198 ***
Correlations Between University-Coded Variables and Interviewer Ratings	
Mother's Self-Consciousness and	
Mother's Comfort Level While Reading	0.237 ***
Mother's Comfort Level During Wheels Task	0.226 ***
Mother's Ease of Ideas to Keep Wheels Task Going	0.189 ***
Mother's Confidence	0.166 ***
Child's Experience of the Session	0.098
Child's Self-Consciousness and	
Mother's Comfort Level While Reading	0.174 ***
Mother's Comfort Level During Wheels Task	0.062
Mother's Ease of Ideas to Keep Wheels Task Going	0.048
Mother's Confidence	0.105
Child's Experience of the Session	0.218 ***

(continued)

Table 12.12 (continued)

Measure	Correlation[a]
Correlations Between Maternal Self-Report of Subjective Well-Being and Other Measures	
Mother's Comfort Level While Reading	0.300 ***
Mother's Comfort Level During Wheels Task	0.259 ***
Mother's Ease of Ideas to Keep Wheels Task Going	0.230 ***
Mother's Confidence	0.332 ***
Child's Experience of the Session	0.362 ***
Mother's Self-Consciousness	0.237 ***
Child's Self-Consciousness	0.154 **
Sample size	290

Sources: New Chance coded observational study variables, variables from brief interview accompanying observational session, and observational study interviewer ratings.

Notes: Calculations for this table used data for all 290 respondents for whom there were 18-month follow-up survey data and a completed observational session, including those with values of zero for outcomes and New Chance enrollees (i.e., experimentals) who did not participate in the program. The sample sizes may fall slightly short of this number because of missing data.

[a] Pearson product-moment correlations. Statistical significance levels (i.e., the probability that the observed relationship is due to chance) are indicated as *** <= 1 percent and ** <= 5 percent.

finding that all of the university-coded variables clustered together — and with the mothers' self-reports — is intriguing in raising the possibility of a source effect, it may be that the construct underlying the interviewers' ratings of self-consciousness is sufficiently different from the constructs underlying the other measures to account for the factor pattern.

IV. Determinants of the Session Experience

No doubt the quality of the participants' experience of the session had several determinants, possibly including enduring qualities of the participants, stressors such as the age of the child or distractions during the session, and the performance of the data collection team.

The premise of the observational work is that a challenging, but neutral, stimulus situation can be presented to the dyads that will allow them to imprint their own style of interaction on it. This premise would predict the kinds of individual differences in confidence or comfort during the tasks, for example, that were noted above. In addition, because the teaching tasks were designed to be challenging so that the child would need the mother's help, the activities could involve a challenge that the dyads might perceive as difficult or even stressful. Further, the level of stress might be heightened by limitations related to the child's age or by occurrences that disrupt the interaction or divert the mother's attention. The data collection team, on the other hand, was supposed to present a neutral backdrop and was not supposed to contribute additional stress; rather, it was to make the mother and child feel as comfortable as possible — within the bounds of the study procedures and guidelines — even engaging in "crowd control" to enable the dyad to work without interference. In addition, the interviewer was to structure the stimulus situation appropriately and ensure that the mother understood what she was expected to do, so that she could work as effectively as possible with her child.

As might be expected, the age of the child did play a role in most of the measures of the subjective experience of the session that were rated in the university laboratories or reported by the mothers. Correlations between Child's Age and these measures appear in Table 12.13. For maternal self-reports, Child's Age was significantly correlated with the Negative Experience subscale of Mother's Overall Subjective Well-Being During Session, but not with the Positive Experience subscale. As previously noted, the Negative Experience subscale focuses more directly on stress or difficulties surrounding the tasks, which we would expect to be greater when the children are younger. The strongest relationship ($r = .359$, $p < .01$) was between Child's Age and Child's Experience of the Session, as assessed by the Minnesota coders. This measure describes how well the child "weathered" the challenges of the activities, in terms of the child's feelings of competence and feelings about his or her relationship with the mother. Since the session posed greater challenges and presumably more stress the younger the child, this finding should not be surprising. Similarly, there was also a reasonably strong association between Child's Age and Mother's Ease of Ideas to Keep Wheels Task Going ($r = .289$, $p < .01$), which seemed to reflect the fact that this task could pose a substantial challenge to mothers of younger children, as they tried to develop strategies to elicit the names of objects with wheels. Not surprisingly, Child's Age was unrelated to Mother's Comfort Level While Reading. As the only activity that did not require the mother to teach or the child to perform, it should not have posed any particular age-related challenges. Likewise, Child's Age was unrelated to the interviewer

Table 12.13

Correlations of Child's Age with Subjective Experience Measures

Measure[a]	Correlation with Child's Age[b]
Mother's Overall Subjective Well-Being During Session	0.182 ***
Positive Experience Subscale	0.103
Negative Experience Subscale	0.176 ***
Mother's Comfort Level While Reading	0.015
Mother's Comfort Level During Wheels Task	0.169 ***
Mother's Ease of Ideas to Keep Wheels Task Going	0.289 ***
Mother's Confidence	0.249 ***
Child's Experience of the Session	0.359 ***
Mother's Self-Consciousness	0.034
Child's Self-Consciousness	0.063
Sample size	290

Sources: New Chance coded observational study variables, variables from brief interview accompanying observational session, and observational study interviewer ratings.

Notes: Calculations for this table used data for all 290 respondents for whom there were 18-month follow-up survey data and a completed observational session, including those with values of zero for outcomes and New Chance enrollees (i.e., experimentals) who did not participate in the program. The sample sizes may fall slightly short of this number because of missing data.

[a] For all variables, a higher score indicates a more positive experience (e.g., greater comfort/ease/confidence, less self-consciousness).

[b] Pearson product-moment correlations. Statistical significance levels (i.e., the probability that the observed relationship is due to chance) are indicated as *** <= 1 percent and ** <= 5 percent.

ratings of Mother's Self-Consciousness and Child's Self-Consciousness, which would not necessarily have been affected by the relative difficulty of the activities.

Having noted the variation in interviewer behavior earlier in this chapter, it is natural to wonder what role the interviewers may have played in determining how the session was experienced by the participants. But the possible relationship between interviewer behaviors and the experiential aspects of the session may be very complex. Interviewers were charged not only with helping the dyads to feel comfortable, but also with maintaining the integrity of the protocol. Yet an action that threatens the integrity of the protocol, like making suggestions to the mother, would not necessarily detract from the mother's experience of the session. Similarly, a problem like inappropriate familiarity might actually enhance rapport and comfort in small doses but make the mother feel very uncomfortable in large doses. Yet the coders' commentary rarely contained enough detail to predict the likely effect of the various deviations on the affective experience of the session.[10] For this reason, interviewer proficiency is represented by the overall-Severity of Deviations score discussed earlier, as the only available proxy for the quality of their performance in administering each session.

In addition to the age of the child and interviewer proficiency, other variables might be associated with the participants' subjective experience of the session, including the ethnic match between interviewer and respondent, whether the interviewer was previously known to the respondent from the New Chance 18-month survey, and the number of 18-month survey interviews the interviewer had completed (as a proxy for more general experience with young, disadvantaged women and their children.)[11]

In trying to understand which, if any, of these attributes of the sessions might have contributed to the laboratory, interviewer, and maternal ratings, regression models were estimated, using Interruptions of the Session, Severity of Deviations, three interviewer characteristics (Ethnic Match of Interviewer and Respondent, Interviewer Same as for 18-Month New Chance Interview, and Interviewer's New Chance Experience), and Child's Age as predictor variables. The significant relationships are shown in Table 12.14. As shown in the table, Child's Age appeared to be the predominant determinant of the participants' subjective experience of the session. The regression analysis supports the correlational findings in Table 12.13, even when other potential contributors to the experience are included in the model. Child's Age is significantly

---

[10] An approach considered was to sort the problem types according to affective and cognitive dimensions, that is, to determine whether they should primarily influence the affective experience of the session or whether they should affect the mother's understanding of the task, her concentration, and/or her ability to identify successful strategies for working with her child. In actuality, however, it was not practical to separate the two. A confusing presentation that leaves the mother uncertain about what she is supposed to do may not only deprive her of the cognitive tools she needs to approach the task but may also make her anxious or angry. Similarly, the emotional state engendered by the interviewer's condescension or rudeness may interfere with information processing activities needed to understand the objectives of the task or develop teaching strategies.

[11] In 61.7 percent of the completed cases, interviewers and respondents were of the same ethnicity, and in 59.3 percent, the same interviewer conducted both the New Chance 18-month follow-up interview and the observational session. The 290 respondents were paired with interviewers varying greatly in the amount of experience they had interviewing the young, disadvantaged women who made up the New Chance sample. Interviewers' experience conducting New Chance 18-month interviews ranged from 3 to 158 interviews, with an average across the completions of 37.7, and a standard deviation of 37.9.

Table 12.14

Predictors of Mothers' and Children's Subjective Well-Being During Observational Session

Measure of Well-Being[a]	Predictor	Parameter Estimate[b]
Mother's Overall Subjective Well-Being During Session	Child's age	0.379 ***
Mother's Comfort Level During Wheels Task	Child's age	0.132 ***
Mother's Ease of Ideas to Keep Wheels Task Going	Child's age	0.347 ***
Mother's Confidence	Child's age	0.227 ***
Child's Experience of the Session	Child's age	0.545 ***
Mother's Self-Consciousness	Interviewer's New Chance experience	-0.013 **
Child's Self-Consciousness	Interviewer's New Chance experience	-0.012 **

Sources: MDRC calculations from New Chance observational study survey data, interviewer ratings, and ratings of the taped sessions by the Harvard and Minnesota laboratories.

Notes: Calculations for this table used data for all 290 respondents for whom there were 18-month follow-up survey data and a completed observational session, including those with values of zero for outcomes and New Chance enrollees (i.e., experimentals) who did not participate in the program. The sample sizes may fall slightly short of this number because of missing or unusable items from some sample members' questionnaires and videotape problems.

[a] For all measures of well-being, a higher score indicates a more positive experience (e.g., greater comfort/ease/confidence, less self-consciousness).

[b] Statistical significance levels (i.e., the probability that the observed relationship is due to chance) are indicated as *** <= 1 percent and ** <= 5 percent.

related to all of the laboratory variables except Mother's Comfort Level While Reading, as well as to Mother's Overall Subjective Well-Being During Session. As previously noted, this pattern is understandable in that the teaching tasks would be expected to pose a greater challenge for younger children and their mothers than the book reading activity.

As foreshadowed by the factor analysis of the measures reported earlier, the interviewer ratings of Mother's Self-Consciousness and Child's Self-Consciousness behaved differently from the remaining measures. Most notably, Child's Age did not play a role in these ratings. This may help to explain why these measures differ from other indicators of the experience of the session. That is, perhaps unlike the other measures, self-consciousness may be unrelated to the difficulty of the session. These measures of self-consciousness were, however, related to the proxy measure of interviewers' experience with members of this population, Interviewer's New Chance Experience, although these relationships were much weaker than those between Child's Age and other indicators of the experience of the session. With a low rating representing a high degree of self-consciousness, the negative coefficients in the table mean that interviewer experience with this population was associated with somewhat higher ratings of Mother's Self-Consciousness and Child's Self-Consciousness. Unless one postulates a greater sensitivity among the more experienced interviewers, this finding is counterintuitive, since one would expect more experienced interviewers to be able to make the dyads feel *less* self-conscious. Although the possibility of an interviewer effect is intriguing, particularly for measures based on interviewer ratings, the staffing patterns for this study may have led to the confounding of specific interviewers with the characteristics examined.

V. Inside the Observational Session: An Overall Characterization

Conduct of the observational sessions posed challenges in terms of the presence of other people during the sessions, a variety of interruptions and distractions, and less than optimal arrangements for the mothers and children to work together on the tasks. For instance, other people were present in more than two-thirds of the sessions, with other children interrupting the session in one-third of the sessions. One session in every seven was conducted with the dyad seated on the floor, presumably because other suitable workspace was not available. Survey interviewers bore the bulk of the responsibility for orchestrating the sessions, and there was considerable variability in their skill at doing so. Deviations from the intended procedures (noted by coders in 42 percent of cases) ranged from those attributable to ambiguities in the study protocol — and perhaps also to some of its more ambitious requirements — to a small minority (4 and 6 percent, respectively, for the two tasks for which we have coders' observations) that suggested either a total misunderstanding of, or complete disregard for, the principles behind the work. Still, our measures of interviewer performance suffered from some limitations, and the appropriate standard for assessing their performance is not entirely clear. The findings suggest some room for improvement — not surprising in an effort that was breaking new ground.

Interestingly, the age of the child, which could affect the ease with which both the mother and child played out their roles, was also related to the interviewer's performance: the younger the child, the lower the indicators of subjective experience of the session, but also the greater the frequency of interviewers' deviations from the study protocol and guidelines in administering the session. While the various indicators of the experience of the session were associated with one another, the child's age was the common thread that united them; other variables examined in this study played a minimal role. This finding underscores the importance of selecting tasks for the target age group that pose sufficient challenge to require the mother's help, while being potentially within reach of the child when given this help. At the same time, it suggests defining the target age group for a set of tasks narrowly enough to provide a roughly comparable experience for those at both ends of the age range.

This chapter described and assessed the way in which the observational sessions were actually administered, based on empirical data. In the next chapter we make recommendations for future work based on these findings, and in the final chapter we consider the broader implications of our methodological reflection.

Chapter 13

An Assessment of the Data Collection Effort and Lessons for Future Research Efforts

Carolyn A. Eldred

In this chapter we evaluate the success of the observational study and offer specific suggestions for future work. We ask whether it is feasible to conduct observational research within a contracted survey research model and answer with a qualified "yes." The chapter discusses the aspects of the approach that we would not change, most notably provision of a standardized script for administering the protocol. At the same time, it recommends that future efforts recognize the central role played by interviewers in such work and the complexity of the task they are being asked to perform. In turn, this recognition implies attention to both the design of data collection protocols and the selection, training, and supervision of interviewers. We argue for an approach built on a recognition of the cognitive demands of the work, augmented by a process for formal certification and ongoing systematic quality control and feedback. While these steps should go a long way toward minimizing systematic interviewer effects, we further argue for controlling for possible interviewer effects in studies employing an experimental design by maintaining approximately the same mix of experimental and control subjects in each interviewer's caseload.

Conducting observational research within a survey framework was a new experience for the observational studies team and is also reasonably unfamiliar to the larger research community. Design and implementation decisions on this study were informed by substantive issues; experience conducting observational research in the laboratory; experience conducting survey research and concern with issues specific to survey measurement; and practical considerations. How did it work out? What lessons can we learn from the effort?

In assessing the work, much of the discussion that follows returns to three themes raised in Chapter 10: the design of data collection instruments, the role of survey interviewers (and their "interface" with the data collection instruments), and working within a survey environment. We begin by discussing those aspects of our approach that we would retain if we were to repeat this study and then consider modifications that build on this work.

I. Aspects of the Approach That Served the Study Well

Chapter 10 discussed the issue of standardization in surveys, and Chapter 11 described the steps taken to attempt to program interviewer behavior, through both the observational script itself and associated instructions for handling situations that could not be programmed explicitly. During the design phase, there had been some concern that the survey approach to scripting the session might constrain the mother-child interaction and prevent the uniqueness of each dyad from emerging. During pretesting and in the study itself, it became apparent that this was not a problem: the protocol elicited considerable variability from case to case. The actual problems presented by standardization were that interviewers could appear wooden in presenting the instructions or the activities or that they could seem inflexible in responding to contingencies that might arise during the session. We believe that these problems can be handled through refinements in training and by ensuring that interviewers have sufficient practice with the instrument, rather than by forsaking standardization.

However, a greater problem was that interviewers departed from the standardized *principles* we attempted to instill, particularly in dealing with issues outside the scripted data collection instrument (for example, answering the mother's questions, identifying situations warranting intervention). This experience is consistent with survey findings that interviewer variation and effects are most common in situations that require interviewers to exercise judgment or use discretion (as in probing to obtain answers to open-ended questions) (Groves and Magilavy, 1986; Groves, 1987). Thus, we would not want to retreat from standardization, which seems necessary if survey interviewers are used to collect the data. In fact, we would want to do a better job of ensuring a standardized performance.

The training materials that accompanied the data collection instrument turned out to be comprehensive and well conceived, anticipating the range of situations that interviewers actually encountered in the field. The good and bad news is that virtually all of the interviewer errors observed on videotape and requiring feedback to the field once the study was under way represented deviations from these reasonably detailed study specifications rather than unanticipated situations. We will return to this issue.

The interviewers were generally successful at helping the mothers to feel comfortable during the session. Of course, these families had recent prior experience as survey participants, which may have helped. Despite variation in interviewer performance, experience, and background characteristics, it is encouraging that the primary influence on the mother's subjective well-being during the session (based on several different measures from several different sources) was the age of her child. This is the one legitimate factor expected to contribute to the stress associated with the session, which, other things being equal, should be more challenging with younger children.

Pretesting had allowed us to anticipate the complexities of collecting data in the homes of low-income families, including the possibility of other adults and children being present, lack of privacy to explain the tasks to the mother without the focal child present, noise, and lack of an ideal workspace. Our strategies for handling such complexities worked reasonably well, but had occasional unintended consequences (for instance, technical quality suffering because the videographer was entertaining the children, or the focal child continuing to be interested in the videographer's toy after starting to work with the mother).

Finally, we were rewarded for the attention paid to technical issues (related to taping) during the design phase. In general, the technical quality of the tapes was quite good, with lapses probably attributable to the need for videographers to divide their time between technical responsibilities and "crowd control" and as a result failing to notice, for example, when a tape was picking up excessive background noise. Decisions to employ an auxiliary light and microphone and to cover the work surface with a light-colored fabric all seem especially prudent in retrospect. In addition, creating audiotapes as well as videotapes was valuable for the Harvard team's transcription and coding of the book reading and wheels tasks.

II. Recommended Refinements to the Approach

Because the role of the interviewer is central to this work, most of the ensuing discussion is related in some way to interviewers — the tasks required of them and their proficiency at performing these tasks. Some recommendations for refining the approach pertain to the design of the protocol and script for administering it, to make the interface between data collection instrument and interviewer function more effectively. In designing such a study in the future, we would focus carefully on the potential cognitive and situational demands on interviewers while in the home and do everything possible to make their task more straightforward, so that a sound, standardized delivery would be the uniform result. In particular, to the extent possible it would be desirable to minimize the need for interviewers to exercise judgment or discretion. Other refinements would include increased attention to interviewer selection, training, credentialing, and management.

A. A Framework for Conceptualizing the Interviewer's Role

A cognitive perspective is useful for viewing the lay interviewer's assignment. Such a perspective suggests a set of questions that are not routinely raised in survey practice about the way interviewers define and perform their assignments. Such questions could fruitfully be considered during each phase of a challenging survey-based effort such as this one — when overall design decisions are being made; when instruments and procedures are being designed; when interviewers are being selected, trained, and supervised; and during analysis, when an awareness of measurement complexities can help in assessing the degree of confidence to place in specific measures. Questions about interviewers cluster into two sets of issues: their conceptualization of the "mission" they are asked to perform and their response to the cognitive demands presented by a specific assignment.

With respect to conceptualization of the task, we could speculate that the way in which interviewers define their mission in assuming these new responsibilities will play a role in how, and how well, they handle them. For example, as interviewers become providers of data by making substantive ratings, how much importance do they attach to this role, and to what extent do they differentiate these ratings from the routine ratings (of respondent cooperativeness, for example) usually made at the end of an interview but not intended as actual measures to be used in the analysis? To what extent do they understand and internalize the message that they are supposed to adhere to standardized criteria rather than offer their own opinion? (After all, the kinds of observations they are asked to make are much like the "opinions" people routinely offer in their daily lives — for example, the child is affectionate toward his mother, the house is dirty.) In administering an observational protocol, how clearly do interviewers perceive that it is different from a survey interview, in terms of ceding control of the unfolding interactions to the dyad while they are under way, but similar in its requirements for neutrality and a nondirective stance?

Training and explicit discussion of the issues above may help to ensure that interviewers understand and accept their role in administering an observational session and/or in providing substantive data. But there is also room for attention to the cognitive demands of the assignment that interviewers are asked to perform. Does the assignment overload them with the need to attend to too many things at once? Does it require them to make overly subtle distinctions? For example, what are the cognitive demands on an interviewer expected to make "incidental" ratings while performing other activities in the home? How well can interviewers monitor mother-child interactions while administering a questionnaire or observational protocol? In making substantive ratings, to what extent is it possible to overcome interviewers' idiosyncratic thresholds and response styles to reflect the researchers' intent? (The larger survey literature indicates that respondents bring different "response sets" to an interview; there is little reason to expect interviewers to be immune to such tendencies, although it is hoped that such response sets can be attenuated through training.) How do interviewers approach related but conceptually distinct parenting measures? What implicit basis of comparison do interviewers use in making ratings, and how is this affected by the ethnicity and socioeconomic status of interviewer and respondent? To what extent do interviewers comprehend the general principles prescribed for handling the specific situations that may arise during an observational session, and what cognitive processes do they go through in applying them?

B. Specific Recommendations for Future Work

Perhaps as foreshadowed by the discussion above, there were some specific aspects of the data collection instrument that seemed to present particular difficulties for interviewers. For instance, it was difficult for them to determine whether or not the requested discussion had occurred after the book reading and then to react appropriately. If anticipated, such a problem might have been handled by providing specific examples of when the "please discuss the book" prompt was to be used or by developing a generic prompt that could be used regardless of the nature or amount of discussion that had occurred. Another example of a tool that might have been helpful to interviewers would have been some way of confirming the mother's understanding of the instructions to the wheels task. As previously noted, in several instances the mother did not perform the task as instructed, but it was unclear whether she had understood the instructions in the first place. Another common problem that became apparent when viewing the tapes was that there was considerable variation from case to case in the time allotted to each task, coupled with a perception in some cases that the interviewers were not handling the timing of tasks very well (for example, by allowing tasks to go on too long or terminating them abruptly). While interviewers were instructed to record the start and expected end time for each task, timing was not emphasized and they were not provided stopwatches. We had not wanted to make the participants feel self-conscious by emphasizing timing, but in future work we would recommend providing the interviewers with stopwatches (or at least watches that measure seconds) and trying to standardize the time spent on each task across the sample. (In subsequent work we have found that mothers seem to understand when they are told that their performance is not being timed but that the interviewer simply wants to make sure the session does not last too long.) Finally, if at all possible, we would want to design tasks that are not very sensitive to the workspace used or the physical arrangement of mother and child and would particularly like to avoid a requirement that the dyads be rearranged part way through the session. In designing the protocol for the JOBS Observational Study (Wave 2), which is currently being conducted by the New Chance Observational Study team, we took such lessons to heart.

But instrument design is only one part of the equation. More attention should be paid to interviewers themselves: their selection, training, performance evaluation, supervision, and assignment of cases. Although many survey researchers are beginning to call for greater attention to interviewers, such attention is even more critical for a study like this one, which demands so much of them. Because of its challenges, an observational study may require more aggressive management than traditional surveys. In terms of project staffing, it should not be assumed that every interviewer can master the work. At minimum, such work requires an interviewer who is willing and able to adhere to the study specifications and incorporate supervisory feedback to improve her (or his) performance. Survey organizations fielding such a study might start by identifying interviewers whose previous work suggests that they are likely candidates for such a study, and possibly even "auditioning" them for the work. But they should also start with the assumption that staffing cannot be finalized until each interviewer has demonstrated the ability to do this work.

Interviewer training is another key area we would seek to enhance. Even if the data collection instrument can be kept reasonably straightforward, it would still not be feasible or desirable to attempt to script every contingency that may arise during the session. Thus, interviewers must learn the general principles they will need to apply during the observational interaction, and they must also learn how to label what is going on so that they can apply the appropriate principle. Since some interviewers did master these lessons pretty well, it may be that interviewer selection (and performance evaluation) will go a long way toward addressing this issue. Indeed, those interviewers who volunteered unsolicited advice to the mothers were not struggling with subtle distinctions but either missed the point entirely, went into the field without adequate preparation or review, or were insufficiently committed to following the protocol.

We believe, however, that some refinements of the training approach could be helpful. Thinking about the interviewer's job as a series of cognitive tasks suggests some possibilities. Consider the need for interviewers to master the following principle: "When the mother asks a question to understand the objective of a task or to clarify the instructions, answer her question directly and clearly. But when the mother asks a question about how to work with her child, deflect the question, as in 'do whatever you think is best' or 'do whatever you and your child usually do together.'" Since quickly labeling or categorizing the mother's request and offering the appropriate response can present a challenge, perhaps this task could be broken down into two steps for training purposes. Thus, interviewers might practice labeling different scenarios as either a request for clarification or a request for direction. Once they master that distinction, they could formulate the appropriate response, which is probably the easier part of the task. Such a focus on a particular principle during training would also emphasize how important it is to the research. The interviewer's overall responsibility for orchestrating the session might be broken down into components for training purposes in other ways as well. Exercises might be developed to expand interviewers' repertoires for dealing with the specific demands of administering an observational protocol. For example, interviewers especially need to remember to sit back and let the interaction unfold while each task is under way, rather than feeling personally responsible for everything that is occurring. Interviewers might also practice the stock phrases that they will be expected to use (for example, to reinforce effort rather than performance), so that they become as second nature to them as the stock phrases they use in survey interviewing (for example, when probing unclear responses). In sum, interviewer training would be designed around explicit recognition of the cognitive and practical challenges of this work, to give interviewers the various tools they will need to administer the session well. Our team followed such an approach in training interviewers for the JOBS Observational Study (Wave 2).

Formal evaluation of each interviewer's performance should also be part of the process. This assessment can begin during training, which can provide an initial screening mechanism.

Before interviewers begin work with research sample members, their performance should be assessed in practice interviews in the field. Such practice interviews, interspersed with supervisory feedback, should quickly lead either to a high level of proficiency or dismissal from the study. In the New Chance Observational Study, we were troubled by finding that the critiques we fed back to interviewers basically restated material in the interviewer's manual and covered in training. We were also troubled by the lack of an effective mechanism for communicating these critiques. We have since developed a checklist specifying those aspects of performance essential to proper administration of the study and representing the various kinds of errors observed on the tapes. This checklist should provide a more effective ongoing reference document about conducting the session than the more elaborate text in the interviewer's manual. It also communicates to interviewers exactly how they will be evaluated. Finally, with commentary on errors in administration, it becomes the evaluation instrument itself, to be used for initial certification of interviewers, periodic assessments during the study, and tracking of interviewer performance for quality control purposes.[1]

The field effort should be managed in a way that incorporates interviewer performance evaluation and quality control into ongoing survey operations. Because this type of work remains unusual in a survey environment and presents special challenges, it may be tempting to regard it as something exotic requiring the involvement of staff with an academic research background, but outside the field operation. But if interviewers are expected to master guiding principles and become proficient at administering observational sessions, surely those who train and supervise them should be expected to differentiate good performance from bad and to take appropriate actions based on such assessments. While those who design and/or commission the research will undoubtedly want to review the work at various points in the study, ongoing quality control needs to be located where it can be incorporated into weekly supervisory discussions, staffing decisions, and so forth. A production orientation will always feature prominently in survey supervision and staffing, but issues of quality can play an equally prominent role in management of the field work.

Other issues in managing the field work relate to the way in which cases are assigned to interviewers. The approach to certification, quality control, and survey management described above should go a long way toward ensuring fairly uniform, and uniformly proficient, performance on the part of interviewers. This, in turn, should limit the variance associated with interviewers. However, since we still know little about the potential for interviewer effects in an observational session (or in connection with interviewer ratings or the kind of open-ended time use question used in this study), we should strive for a more even distribution of cases across interviewers than was the case in this study. At minimum, for studies based on an experimental design, each interviewer's caseload should be monitored to make sure that the breakdown of experimental and control group members roughly parallels that of the overall sample. MDRC is implementing this recommendation in several surveys.

Of course, practical considerations such as cost may influence decisions about embracing the refinements discussed here or others that may seem promising. The idea of breaking interviewer training into discrete cognitive tasks should have minimal cost implications. The initial cost implications of the proposed interviewer selection and certification procedures are probably greater, but they could be compensated by lower attrition over the course of the study if a relatively constant core of committed, proficient interviewers results. Other possible steps, for example, those that would constrain decisions about staffing and assignments, could have the effect of limiting some of the flexibility that "gets the job done" and so might affect schedule or response rates, if not cost. Thus, while we would surely wish to make some changes in the design and procedures if we were to repeat this study, any study like this would continue to be faced with tradeoffs among competing priorities. Nonetheless, the need to maintain an appropriate balance of experimental and control subjects in an interviewer's caseload is sufficiently important in research employing an experimental design that it should become routine practice in any case.

---

[1] The evaluation tool developed as a result of the New Chance Observational Study and now in use in the JOBS Observational Study (Wave 2) is available upon request from MDRC.

III. The Feasibility of Observational Research Within a Survey Framework

In concluding these reflections on our experience in conducting the New Chance Observational Study, we return to the question of whether it is feasible to carry out such work within a contracted survey research model. Our answer is a qualified "yes." To optimize the value of the data collected through such methods, however, it is necessary to recognize the complexity of the interviewer's assignment and to pay careful attention both to the design of data collection protocols and to the selection, training, and supervision of interviewers. We have argued for an approach built on a recognition of the cognitive demands of this work, augmented by a formal interviewer certification process, with ongoing quality control and feedback to interviewers on their performance. With such measures in place, we believe that observational research within a survey framework is a viable methodological option for developmentalists.

In this chapter we evaluated the success of the observational study and offered specific recommendations for future work. In the final chapter we consider the broader measurement implications of the work.

Chapter 14

Measurement Implications of the New Chance Observational Study

Carolyn A. Eldred

This chapter concludes Part II with an assessment of the methodological implications of the New Chance Observational Study and of our reflections on it. It discusses the process of deciding whether and how to undertake observational work within a survey research context, first focusing on methodological considerations such as feasibility and data quality and then turning to substantive issues, specifically the relative contribution or "value added" of observational and other measures. The design of specific measures, instruments, and procedures is also discussed, in the context of a belief that efforts to examine parenting and child outcomes using survey methods must reflect an understanding of what the survey model is and is not. Conceptualizing the roles of both respondent and interviewer during a survey interaction (including observational work) in terms of their cognitive demands is helpful in integrating what is known about survey response, assessing what is likely to be possible, and identifying the steps necessary to implement a given component of a design. In addition, controlling for systematic interviewer effects in studies employing an experimental design is critical.

The various measures of parenting employed in this study represented an unusual array of sources, methods, perspectives, and rigor, suggesting a valuable opportunity for scrutiny of the methods involved in producing them. Our reflection on this challenging effort is intended both to inform our own future work and to serve as a bridge to future endeavors by our colleagues in the child development, survey research, and policy and evaluation research communities.

Part II began with a discussion of the implications of what it means to do research within a survey framework. This discussion was intended in part to establish realistic expectations about the survey method and also to "raise the consciousness" of survey researchers themselves about the measurement challenges they need to address in undertaking innovative survey efforts. Chapter 10 also looked to the somewhat fragmented literature on survey measurement, attempting to extract and integrate what is known in a way that would help in conceptualizing the current effort. In subsequent chapters we provided specific information about how this work was carried out, empirical data bearing on the success of the effort, and recommendations for future work as a guide to those contemplating such an effort. With this information as a backdrop, we conclude this volume by examining the methodological or measurement implications of the work.

In designing a study, there are many decisions to be made, such as whether to turn to a survey model at all and for which measurement tasks. Because these choices must be made within the constraints of available resources, it is critical to make decisions that optimize the value of the information collected. For instance, survey-based observational research could be expected to leverage larger, more representative, and/or less accessible samples than laboratory-based research, but it is much more expensive and challenging than survey work involving only traditional interviews. Yet relying solely on maternal self-reports, even supplemented by interviewer ratings, may not tell the whole story. How then to choose?

Decisions need to be made, first about the overall design of the work (for example, whether to embed an observational study in a larger survey effort) and also about the design of specific measures, instruments, and procedures. These two interdependent decisions should not be made sequentially, however, because knowledge of what it takes to employ various methods and what can be expected in return should inform overall design decisions. As we have seen, especially when the effort presents particular challenges, attention to measurement processes is critical.

I. Overall Study Design

The monograph has maintained a dual focus on substantive findings and methodological issues, and these two subjects continue to provide the framework for thinking about study design.[1] Thus, the methodological assessment focuses on feasibility and data quality, while the substantive assessment focuses on the relative contribution of each potential component of the work to its substantive goals. When more challenging or costly components are considered, the latter becomes an issue of the *added* value of the component.

A. Methodological Considerations: Feasibility and Data Quality

With a few caveats, we believe that what we attempted to do in measuring parenting and child outcomes among New Chance sample members is feasible. While perhaps self-evident, we think that this in an important conclusion. We have found it feasible to obtain maternal self-reports on parenting, obtain interviewer ratings, administer cognitive assessments to children, and administer observational protocols, all within a survey framework. In other work, interviewers have actually interviewed children, so this appears to be a feasible option as well.

Clearly, some questions remain, however, and there is more work to be done. For instance, although interrater reliabilities have been established for the original form of the HOME, we do not know of any comparable information about the short form used in survey work.[2] Since it is generally not practical to assess interrater reliabilities in the course of survey data collection without mounting special methodological studies, we know little about this aspect of the "quality" of interviewer ratings or the extent of interviewer effects. In the current work, we did find some evidence of a small, but significant, interviewer effect for ratings of self-consciousness (the only interviewer ratings examined from a methodological perspective), so we continue to urge attention to interviewers, especially in nontraditional measurement situations where their potential influence on the data is most likely. Ideally, studies would be undertaken that randomly assign cases to interviewers as a way of studying interviewer effects. At minimum, however, for studies employing experimental designs, interviewer assignments should be monitored to ensure that they are balanced between experimental and control cases, so that experimental impacts are not confounded with interviewer effects. In Chapter 13, we discussed numerous specific design and operational refinements that could enhance survey-based observational work. We also introduced a conceptual framework — the cognitive approach — that offers an integrated approach for thinking about the design of complex survey endeavors. We return to this theme at the conclusion of the chapter.

---

[1] Resource constraints and the need to make optimal choices are assumed as the backdrop but are not explicitly addressed here

[2] A team of researchers headed by Zaslow at Child Trends will be examining interrater reliability as well as test-retest reliability for a modification of the short form of the HOME, as part of an ongoing project focusing on survey measures of parenting, under an NICHD grant.

B. Substantive Considerations: The Relative Contribution or "Value Added" of Measures Based on Various Methods

As previously noted, the study included a rich array of survey-based parenting measures. A basic question underlying our methodological reflection concerns the "value added" of the observational work, above and beyond parenting measures obtained through maternal self-report and interviewer observations. Given the challenge and cost of embedding observational work within a survey effort, this is a critical question. As discussed in Chapter 7, our conclusion is that the survey and observational parenting measures examined in this study were complementary rather than duplicative — related, but not overlapping. Further, the observational measures added to our ability to predict developmental outcomes. The observational work appeared to open a window on the qualitative aspects of parenting, providing an opportunity for rigorous assessment of important constructs based on observation of the mother-child relationship during structured activities. Perhaps this should not be surprising, since we can assume that there are both cognitive limits and limits posed by social desirability response sets, for example, on what mothers can report about their own parenting styles and behaviors. But observational measures are not a substitute for more traditional survey measures. For example, mothers are the only ones who can speak to a construct like stress in parenting, underscoring the importance of self-report data. At the same time, interviewers who spend time in a respondent's home have access to a different set of information about the home environment and mother-child interaction than that captured on videotape during structured activities, so their perspective may be important as well.

Survey-based observational work could be undertaken as part of a freestanding study, in which all participants provide both self-reported data, perhaps enhanced by interviewer ratings, and observational data. Or, observational work could be embedded within a larger survey effort, as was the case in this study. In considering the latter approach, part of the design work should focus on the implications of this strategy. Thus, while we have already noted the *substantive* value added of observational measures, we now suggest a focus on the *analytical* contribution of obtaining such measures for a subset of a larger sample. Other things being equal, the ability to leverage from a subsample will be greater when the subsample is drawn randomly from the larger sample (which was not possible in this study).

The meaning of "analytical leverage" may vary from study to study. In some cases it may be appropriate simply to generalize to the larger sample and/or to the population from which it was drawn, through statistical weighting procedures. Or, it might be desirable to draw on the richness of the observational measures to help in interpreting data from other sources. For instance, observational measures could help in determining how much confidence to place in other measures that may be subject to different potential method effects. Measures based on mother-child interaction coded under rigorous circumstances may also help in teasing out findings from other sources that are inconclusive or seem counterintuitive. In addition, observational measures from a subsample may be viewed in the same way as the post-survey focus groups that are sometimes undertaken to "put flesh on the bones" of quantitative survey findings. These examples for leveraging observational data argue for thinking about the potential analytical contribution of such measures — in addition to the substantive added value of the measures themselves — during study design.

To the extent that any of our conclusions about the relative contribution of various measures are specific to this study, they could vary with the particular parenting constructs under consideration, the conceptualization of these constructs (for example, broad versus focused), the specific measures and measurement processes contemplated, and the age of the children to be studied. Thus, in addition to our specific conclusions, we offer a more general approach for conceptualizing study design and associated measures. In assessing the potential substantive contribution of measures derived from a variety of methods and sources, we would suggest mapping them according to a framework like that in Table 7.2 (Chapter 7). In this way the potential strengths and limitations of each become clearer, making it possible to make informed choices (within resource constraints) and spread the measurement risk across a "balanced portfolio" of methods and measures.

II. The Design of Measures, Instruments, and Procedures

Regardless of the mix of methods and measures used, efforts to examine parenting and child outcomes using a survey model must reflect an understanding of what the survey model is and is not. The need to keep the survey model keenly in mind applies not only to developmentalists (or others who commission survey research) but to survey researchers themselves, who will inevitably play a role in reconciling the demands of innovative efforts with survey traditions. In addition, it is important for all to recognize the challenge of some of the more innovative or demanding work. The fact that a researcher can write a question does not necessarily mean that a survey respondent (or an interviewer) can answer it. The fact that a laboratory-based observational protocol can be adapted for survey administration does not guarantee that it can, or will, be administered in the field as the researchers intend. Sound design of measures, instruments, and data collection procedures is important. But so are sound quality control and survey management procedures, as a partial antidote to the production emphasis of survey work.

Certainly design efforts should build on what is known about survey response, to the extent that the literature holds germane lessons. But survey design also needs a way of thinking about what it is reasonable to ask respondents and/or interviewers to do. For instance, few would expect a mother to be able to report how frequently she uses "decontextualized language" with her child. Less clear, however, is how well a survey interviewer can be expected to observe, judge, and recall how frequently a particular behavior takes place (for example, the child showing anger or hostility toward the mother), during an interview in which she is also asking questions and recording answers.

As previously discussed, conceptualizing the roles of both respondent and interviewer during a survey interaction in terms of their cognitive demands is helpful in integrating what is known about survey response, assessing what is likely to be possible, and identifying the steps likely to be required to implement a given component of a design. Indeed, Dippo and Norwood

(1992) have recently observed that the cognitive model holds as yet largely untapped promise for thinking about interviewers and the tasks they face during an interview. Furthermore, the cognitive model can be extended to situations that place somewhat different demands on interviewers, such as the need to make substantive ratings or administer an observational protocol. The cognitive processes employed by interviewers in responding to unprogrammed situations is of particular interest. In fact, the field of survey research is currently moving in the direction of extending the cognitive model to interviewers as well as respondents.

Given the central role assigned to survey interviewers in a study like this, and the variation in their ability to perform that role well, the cognitive model offers considerable benefits in assessing and accommodating the cognitive processes involved in data collection. A clearer focus on the cognitive demands facing both respondents and interviewers during a given survey interaction should make it possible to design a more effective interface between the two, namely, the data collection instrument and associated procedures. At the same time, interviewer effects represent an ever-present source of potential bias despite efforts made to minimize them. To *measure* the extent of such effects, random assignment of cases to interviewers is necessary. And to *control* for such effects in a study employing an experimental design, each interviewer's caseload should mirror the distribution of experimentals and controls in the overall sample.

III. A Bridge to Future Work

The New Chance Observational Study employed a multifaceted measurement strategy, involving a variety of data sources and methods. As such, it provided a unique opportunity for methodological reflection. Because Part II of this monograph represents an attempt to distill lessons from our experience and to look ahead to future work having related substantive agendas and similar methodological requirements, it has tended to focus on measurement "vulnerabilities" and ways to address them. But rather than closing on the cautionary note that may have been suggested by previous discussion, we end on a positive, optimistic chord. The New Chance Observational Study represented a challenging, ground-breaking research collaboration, and the study team is grateful for the opportunity to have undertaken it. The study has made important substantive and methodological contributions. Thus, despite the concerns raised in earlier chapters about data collection carried out through a contracted survey research model, there was good variability in the behavioral interactions of the study subjects and meaningful information on the videotapes. And, despite the concern about survey interviewers expressed in Part II, the experimental findings reported in Part I should be resilient in the face of possible interviewer "noise" in the measures, since there were not any systematic experimental-control differences in the interviewers' assignments. From a methodological standpoint, both the accounting of our experience and the opportunity to synthesize and apply the body of knowledge on survey measurement to the issues faced in this work represent a unique contribution that should be helpful to those considering endeavors similar to this study.

Like the New Chance Observational Study, most future survey-based research on parenting will be undertaken primarily to answer substantive, rather than methodological, questions. But the survey method may also provide developmentalists an opportunity to examine and refine their measures and survey researchers an opportunity to improve their methods — especially when given the chance to capture actual interviewer performance on videotape.

Appendix A

Table A.1

Regression Coefficients from a Regression of the Experimental Dummy on Selected Baseline Characteristics

Variable	Parameter Estimate	Standard Error	p[a]
Intercept	0.635 ***	0.028	0.000
Bronx	0.044	0.148	0.764
Detroit	-0.083	0.146	0.571
Harlem	-0.077	0.144	0.596
Lexington, Kentucky	-0.044	0.154	0.776
Philadelphia	-0.081	0.138	0.558
Portland, Oregon	0.141	0.155	0.364
Age 20-22 at baseline	-0.106	0.138	0.443
More than one child at baseline	-0.040	0.090	0.660
Age 16 or younger when had first child	-0.050	0.077	0.518
Ever had an abortion	-0.090	0.080	0.258
Driver's license	0.240 **	0.106	0.024
Ever had a miscarriage at baseline	0.076	0.102	0.455
No home phone at baseline	-0.085	0.087	0.333
Ethnicity black, non-Hispanic	-0.121	0.102	0.236
Highest grade completed 10th or above	-0.059	0.074	0.422
Had a high school diploma or GED at baseline	-0.206	0.160	0.201
TABE level 8th grade or below[b]	0.068	0.064	0.296
Ever had vocational training at baseline	-0.079	0.076	0.301
Highest educational goal: high school	-0.026	0.068	0.703
Received child support at baseline	0.063	0.068	0.354
Used no birth control at last intercourse	-0.029	0.068	0.676
Had regular child care at baseline	0.043	0.067	0.526
Was pregnant three times or more prior to baseline	0.050	0.096	0.602
Youngest child is older than age 1	0.058	0.075	0.443
Scored less than 21 on control scale	0.008	0.074	0.909
Scored less than 35 on self-esteem scale	-0.004	0.082	0.966
Ever repeated a grade	0.000	0.071	0.999
Not receiving AFDC in own name at baseline	0.004	0.088	0.968
Heard about NC in welfare office	-0.071	0.085	0.401
Not JOBS-mandatory	0.060	0.136	0.659
At risk of depression	0.056	0.068	0.412
Currently/previously married	0.149	0.144	0.300
Never on welfare when young	0.088	0.071	0.216
In public or subsidized housing at baseline	0.051	0.076	0.500
Received family planning services before baseline	0.177 **	0.075	0.019
Father of child never sees child	-0.032	0.073	0.662
Does not expect more children	-0.062	0.069	0.369
Lived with own father when 14 years old	0.086	0.079	0.276
Ever employed in past 12 months	-0.082	0.084	0.329
Earned $501 or more in past 12 months	0.071	0.098	0.472
Never employed	-0.103	0.087	0.238
Sample size	290		
R^2	0.140		
F-statistic	0.981		0.509

Source: New Chance baseline enrollment data.

Notes: [a]Statistical significance levels are indicated as *** = 1 percent; ** = 5 percent; * = 10 percent.

[b]The test used to measure reading ability was the reading part of the Tests of Adult Basic Education (TABE). Most sites administered the survey form of the test, but some administered the full reading test.

Appendix B

Table B.1

Observational Study Respondents and Initial Pool of Respondents in the Seven Study Sites

Baseline Characteristics	Observational Study Respondents	Eligible Nonrespondents	Difference	p[a]
Mother's age at enrollment (%)				
16 or 17	19.5	14.9	4.6	0.205
18	24.6	17.7	6.9	
19	25.9	27.0	-1.0	
20 or over	30.0	40.4	-10.5	
Mean age of mother in years	18.8	19.1	-0.3	0.443
Race/ethnicity (%)				
Black, non-Hispanic	83.8	80.1	3.7	0.340
White, non-Hispanic	16.2	19.9	-3.7	
Ever married (%)	5.4	7.9	-2.5	0.321
Number of children (%)				
1	62.3	61.7	0.6	0.457
2	31.3	31.2	0.1	
3 or more	6.4	7.1	-0.7	
Mean age of mother at first birth	16.6	17.0	-0.4	0.751
Highest grade completed at baseline (%)				
8th or below	10.1	9.2	0.9	0.167
9th	22.9	27.7	-4.8	
10th	30.3	25.5	4.8	
11th or above	36.7	37.6	-0.9	
Any employment at baseline (%)	76.8	73.1	3.7	0.397
Welfare status at baseline (%)				
Receiving - own case	82.4	88.7	-6.2 **	0.031
Receiving - other case	13.9	5.7	8.2	
Not receiving welfare	3.7	5.7	-2.0	
Welfare status as a child (%)				
Never received welfare	29.4	34.8	-5.4	0.372
Sometimes received welfare	49.7	43.5	6.2	
Always received welfare	21.0	21.7	-0.8	
Sample size	297	141		

Source: New Chance baseline enrollment data.

Notes: Only sample members in the observational study sites (Bronx, Detroit, Harlem, Lexington, Pittsburgh, Philadelphia, and Portland) were included. Statistical significance levels are indicated as *** = 1 percent; ** = 5 percent; * = 10 percent.

[a] The column labeled "p" indicates the probability that the difference in baseline characteristics is due to random error.

No statistical tests were done to measure the significance of the differences in outcomes across different samples within the experimental and control groups.

BEST COPY AVAILABLE

Appendix C

Table C.1

Regression Coefficients for Analyses Examining Multiple Influences on Mother-Reported Behavior Problems Index

Variable	Model 1 Control Variables Only	Model 2 (Adds Parenting)	Model 3 (Adds Mother's Subjective Well-Being)	Model 4 (Adds Mother's Human Capital)	Model 5 (Adds Larger Social Context)
Controls					
Experimental/control group	2.78	2.76	2.25	2.36	1.41
Focal child's age	-.12	-.01	-.02	-.02	-.04
Focal child's gender	-1.44	-.94	-1.24	-1.47	-2.13
Race/ethnicity	5.60 **	6.95 ***	4.99 *	4.86 *	4.40
Mother's TABE score	.03	.04 *	.05 **	.06 ***	.05 ***
More than one child	.42	.62	.95	1.18	-1.14
Philadelphia site	-.66	-1.31	-1.61	-2.07	-2.46
Portland site	-2.40	-1.84	-.13	.32	-1.92
Parenting					
HOME Emotional Support		.07	.04	.05	.06
HOME Cognitive Stimulation		-.03	.05	.06	.04
HOME Harsh Discipline (higher =less HD)		-.19 ***	-.12 **	-.11 **	-.10
Maternal Warmth		.03	.03	-.01	-.01
Mother's Harsh Treatment		2.14 **	2.05 **	2.05 **	2.33 **
Mother's Confidence		-2.83 **	-2.10 *	-1.74	-2.03 *
Percentage of Immediate Utterances		.36	4.95	3.97	4.37
Objects/Elicitations		-.84	.16	-.07	.52
Mother's Subjective Well-Being					
Mastery score at 18 months			-.30	-.31	-.27
Parenting Stress score at 18 months			.20 ***	.21 ***	.20 ***
CES-D (depression) score at 18 months			.16	.12	-.06
Life Satisfaction score at 21 months			-.01	-.01	.01

(continued)

Table C.1 (continued)

Variable	Model 1 Control Variables Only	Model 2 (Adds Parenting)	Model 3 (Adds Mother's Subjective Well-Being)	Model 4 (Adds Mother's Human Capital)	Model 5 (Adds Larger Social Context)
Mother's Human Capital					
Trade license at 42 months				-.01	.00
High school diploma/GED at 42 months				-.04 **	.03 *
On AFDC at 42 months				-.01	-.01
Employed at 42 months				-.02	-.02
Total earnings months 1-42				.00	.00
Larger Social Context					
Difficult Life Circumstances (42 months)					2.09 ***
Number of moves (42 months)					.37
Living with a partner (18 months)					.03
No social support (42 months)					.04
Satisfaction with social support (42)					-.40
Number of children in household (42 months)					.99
Number of months focal child in day care/preschool (18 months)					.30 *
Focal child in child care before age 1					.03

Sources: New Chance baseline enrollment data, 18-month follow-up survey, coded observational study variables, and 42-month follow-up survey.

Notes: Results rely on data for the 247 observational study mothers who responded to the Behavior Problems Index at the 42-month follow-up survey. Statistical significance levels are indicated as *** <= 1 percent, ** <= 5 percent, and * <= 10 percent, which refers to the probability that the regression coefficient is larger than zero because of chance alone.

Table C.2

Regression Coefficients for Analyses Examining Multiple Influences on Mother-Reported Positive Behavior Index

Variable	Model 1	Model 2	Model 3	Model 4	Model 5
	Control Variables Only	(Adds Parenting)	(Adds Mother's Subjective Well-Being)	(Adds Mother's Human Capital)	(Adds Larger Social Context)
Controls					
Experimental/control group	-9.37 **	-11.26 ***	-9.84 **	-10.05 **	-11.98 **
Focal child's age	0.37	0.06	0.14	0.15	0.19
Focal child's gender	10.47 ***	9.13 **	9.01 **	9.39 **	9.80 **
Race/ethnicity	-17.30 ***	-18.55 ***	-12.28 *	-12.81 **	-8.43
Mother's TABE score	-0.01	-0.03	-0.05	-0.06	-0.04
More than one child	-11.94 ***	-12.41 ***	-12.91 ***	-13.43 ***	-11.40 **
Philadelphia site	2.49	3.72	5.43	6.41	5.98
Portland site	4.84	6.81	4.46	3.76	7.54
Parenting					
HOME Emotional Support		0.02	0.11	0.07	0.07
HOME Cognitive Stimulation		0.04	-0.11	-0.15	-0.19
HOME Harsh Discipline (higher =less HD)		.30 **	0.15	0.15	0.12
Maternal Warmth		.73 **	.62 *	.70 **	.77 **
Mother's Harsh Treatment		-2.22	-2.2	-2.17	-2.25
Mother's Confidence		8.16 ***	6.37 **	5.76 **	5.26
Percentage of Immediate Utterances		18.05	9.47	10.66	9.67
Objects/Elicitations		7.85	3.56	4.68	2.55
Mother's Subjective Well-Being					
Mastery score at 18 months			0.99	1.02	1.19 *
Parenting Stress score at 18 months			-0.11	-0.13	-0.11
CES-D (depression) score at 18 months			-0.26	-0.15	-0.01
Life Satisfaction score at 21 months			.35 **	0.36	0.22

(continued)

Table C.2 (continued)

Variable	Model 1 Control Variables Only	Model 2 (Adds Parenting)	Model 3 (Adds Mother's Subjective Well-Being)	Model 4 (Adds Mother's Human Capital)	Model 5 (Adds Larger Social Context)
Mother's Human Capital					
Trade license at 42 months				0.02	0.03
High school diploma/GED at 42 months				0.07	0.05
On AFDC at 42 months				0.00	0.01
Employed at 42 months				0.06	0.04
Total earnings months 1-42				0.00	0.00
Larger Social Context					
Difficult Life Circumstances (42 months)					-1.45
Number of moves (42 months)					-1.91 **
Living with a partner (18 months)					-0.01
No social support (42 months)					-.18 **
Satisfaction with social support (42)					1.51
Number of children in household (42 months)					-2.05
Number of months focal child in day care/preschool (18 months)					0.52
Focal child in child care before age 1					0.02

Sources: New Chance baseline enrollment data, 18-month follow-up survey, coded observational study variables, and 42-month follow-up survey.

Notes: Results rely on data for the 246 observational study mothers who responded to the Positive Behavior Index at the 42-month follow-up survey. Statistical significance levels are indicated as *** <= 1 percent, ** <= 5 percent, and * <= 10 percent, which refers to the probability that the regression coefficient is larger than zero because of chance alone.

Table C.3

Regression Coefficients for Analyses Examining Multiple Influences on Teacher-Reported Behavior Problems Index

Variable	Model 1 Control Variables Only	Model 2 (Adds Parenting)	Model 3 (Adds Mother's Subjective Well-Being)	Model 4 (Adds Mother's Human Capital)	Model 5 (Adds Larger Social Context)
Controls					
Experimental/control group	.08	.05	-.30	-.49	-.71
Focal child's age	.18	.24 ***	.25 ***	.31 ***	.28 ***
Focal child's gender	-4.03 ***	-3.90 ***	-4.25 ***	-3.78 ***	-3.81 ***
Race/ethnicity	1.68	2.06	1.45	.88	1.02
Mother's TABE score	-.03 ***	-.03 *	-.03 **	-.03 **	-.03 **
More than one child	.22	-.04	.04	.32	-.58
Philadelphia site	1.43	1.16	.94	1.89	1.63
Portland site	2.87	1.89	2.22	2.37	2.50
Parenting					
HOME Emotional Support		-.04	-.05	-.06	-.07 *
HOME Cognitive Stimulation		-.02	.01	.02	.01
HOME Harsh Discipline (higher =less HD)		.02	.04	.05	.05
Maternal Warmth		.10	.10	.09	.05
Mother's Harsh Treatment		.17	.09	.24	.58
Mother's Confidence		-.44	-.13	.11	-.05
Percentage of Immediate Utterances		-6.40	-5.11	-5.84	-5.03
Objects/Elicitations		-5.08 **	-4.89 **	-5.00 **	-4.26 *
Mother's Subjective Well-Being					
Mastery score at 18 months			.23	.29	.31
Parenting Stress score at 18 months			.01	.00	-.01
CES-D (depression) score at 18 months			.14	.19 **	.15
Life Satisfaction score at 21 months			-.03	-.03	-.05

(continued)

Table C.3 (continued)

Variable	Model 1 Control Variables Only	Model 2 (Adds Parenting)	Model 3 (Adds Mother's Subjective Well-Being)	Model 4 (Adds Mother's Human Capital)	Model 5 (Adds Larger Social Context)
Mother's Human Capital					
Trade license at 42 months				-.04 **	-.04
High school diploma/GED at 42 months				.01	.01
On AFDC at 42 months				-.03	-.02
Employed at 42 months				.02	.02
Total earnings months 1-42				.00	.00
Larger Social Context					
Difficult Life Circumstances (42 months)					.85 ***
Number of moves (42 months)					.07
Living with a partner (18 months)					-.01
No social support (42 months)					.03
Satisfaction with social support (42)					.31
Number of children in household (42 months)					.68
Number of months focal child in day care/preschool (18 months)					.14
Focal child in child care before age 1					.00

Sources: New Chance baseline enrollment data, 18-month follow-up survey, coded observational study variables, and 42-month follow-up survey.

Notes: Results rely on data for the 148 observational study mothers and teachers who responded to the Behavior Problems Index at the 42-month follow-up survey. Statistical significance levels are indicated as *** <= 1 percent, ** < =5 percent, and * < =10 percent, which refers to the probability that the regression coefficient is larger than zero because of chance alone.

Table C.4

Regression Coefficients for Analyses Examining Multiple Influences on Teacher-Reported Positive Behavior Index

Variable	Model 1 Control Variables Only	Model 2 (Adds Parenting)	Model 3 (Adds Mother's Subjective Well-Being)	Model 4 (Adds Mother's Human Capital)	Model 5 (Adds Larger Social Context)
Controls					
Experimental/control group	-.78	-2.48	-2.42	-.26	.14
Focal child's age	-1.61 ***	-2.10 ***	-2.09 ***	-2.52 ***	-2.36 ***
Focal child's gender	29.35 ***	26.78 ***	27.07 ***	25.12 ***	27.84 ***
Race/ethnicity	-2.54	-3.93	-4.05	-.55	5.69
Mother's TABE score	.25 **	.13	.13	.13	.14
More than one child	-8.85	-7.41	7.19	-9.32	1.50
Philadelphia site	4.51	5.35	5.83	-.17	4.55
Portland site	-31.9 **	-24.01	-23.70	-24.65	-23.16
Parenting					
HOME Emotional Support		.10	.10	.15	.24
HOME Cognitive Stimulation		.54	.54	.44	.53
HOME Harsh Discipline (higher =less HD)		-.07	-.08	-.12	-.19
Maternal Warmth		-.11	-.13	-.06	.34
Mother's Harsh Treatment		-4.58	-4.70	-5.61	-8.43
Mother's Confidence		7.99	8.03	6.54	8.32
Percentage of Immediate Utterances		62.64 *	63.30 *	66.95 *	57.36
Objects/Elicitations		37.18 **	36.74 **	40.40 **	29.18
Mother's Subjective Well-Being					
Mastery score at 18 months			-.19	-.65	-.55
Parenting Stress score at 18 months			-.02	.13	.04
CES-D (depression) score at 18 months			.03	-.30	.48
Life Satisfaction score at 21 months			.05	.09	.15

(continued)

Table C.4 (continued)

Variable	Model 1 Control Variables Only	Model 2 (Adds Parenting)	Model 3 (Adds Mother's Subjective Well-Being)	Model 4 (Adds Mother's Human Capital)	Model 5 (Adds Larger Social Context)
Mother's Human Capital					
Trade license at 42 months				.26 **	.26 **
High school diploma/GED at 42 months				-.03	-.06
On AFDC at 42 months				.14	.17
Employed at 42 months				-.16	-.14
Total earnings months 1-42				.00	.00
Larger Social Context					
Difficult Life Circumstances (42 months)					-5.98 **
Number of moves (42 months)					-3.18
Living with a partner (18 months)					.06
No social support (42 months)					-.25
Satisfaction with social support (42)					-.76
Number of children in household (42 months)					-9.41
Number of months focal child in day care/preschool (18 months)					-1.46
Focal child in child care before age 1					-.08

Sources: New Chance baseline enrollment data, 18-month follow-up survey, coded observational study variables, and 42-month follow-up survey.

Notes: Results rely on data for the 152 observational study mothers and teachers who responded to the Positive Behavior Index at the 42-month follow-up survey. Statistical significance levels are indicated as *** <= 1 percent, ** <= 5 percent, and * <= 10 percent, which refers to the probability that the regression coefficient is larger than zero because of chance alone.

Table C.5

Regression Coefficients for Analyses Examining Multiple Influences on the Bracken Basic Concept Scale

Variable	Model 1 Control Variables Only	Model 2 (Adds Parenting)	Model 3 (Adds Mother's Subjective Well-Being)	Model 4 (Adds Mother's Human Capital)	Model 5 (Adds Larger Social Context)
Controls					
Experimental/control group	-.38	-.66	-.63	-.62	-.70
Focal child's age	.04 **	.03	.03	.04	.04
Focal child's gender	.63	.57	.48	.49	.43
Race/ethnicity	.10	-.05	.12	.00	-.29
Mother's TABE score	.10 **	.01	.01	.01	.01
More than one child	-1.13 ***	-1.10 ***	-1.15 ***	-1.15 ***	-.58
Philadelphia site	.18	.08	.03	.05	.24
Portland site	-1.65	-1.40 *	-1.48 **	-1.47 *	-1.72 **
Parenting					
HOME Emotional Support		.03 **	.03 **	.03 **	.03 **
HOME Cognitive Stimulation		.01	.02	.02	.02
HOME Harsh Discipline (higher =less HD)		.01	.01	.01	.01
Maternal Warmth		.04	.04	.03	.03
Mother's Harsh Treatment		.16	.18	.19	.17
Mother's Confidence		.37	.35	.36	.31
Percentage of Immediate Utterances		.69	.87	.81	.74
Objects/Elicitations		.09	.11	.17	.19
Mother's Subjective Well-Being					
Mastery score at 18 months			.04	.04	.03
Parenting Stress score at 18 months			.03	.02	.02
CES-D (depression) score at 18 months			-.01	-.01	-.02
Life Satisfaction score at 21 months			.01	.01	.01

(continued)

Table C.5 (continued)

Variable	Model 1 Control Variables Only	Model 2 (Adds Parenting)	Model 3 (Adds Mother's Subjective Well-Being)	Model 4 (Adds Mother's Human Capital)	Model 5 (Adds Larger Social Context)
Mother's Human Capital					
Trade license at 42 months				.00	.00
High school diploma/GED at 42 months				.00	.00
On AFDC at 42 months				.00	.00
Employed at 42 months				.00	.00
Total earnings months 1-42				.00	.00
Larger Social Context					
Difficult Life Circumstances (42 months)					-.03
Number of moves (42 months)					.04
Living with a partner (18 months)					.01
No social support (42 months)					-.01
Satisfaction with social support (42)					-.06
Number of children in household (42 months)					-.52 *
Number of months focal child in day care/preschool (18 months)					.01
Focal child in child care before age 1					.01

Sources: New Chance baseline enrollment data, 18-month follow-up survey, coded observational study variables, and 42-month follow-up survey.

Notes: Results rely on data for the 252 observational study children who were administered the direct-assessment Bracken measure at the 42-month follow-up survey. Statistical significance levels are indicated as *** <= 1 percent, ** <= 5 percent, and * < =10 percent, which refers to the probability that the regression coefficient is larger than zero because of chance alone.

References

Aber, J. L., Brooks-Gunn, J., and Maynard, R. (1995). Effects of welfare reform on teenage parents and their children. *The Future of Children: Critical issues for children and youths, 5*(2), 53-71.

Aber, J. L., Goodman, G., and Morris, P. (1993, November). The effects of welfare reform on the preschool children of teenage mothers: The mediating role of maternal characteristics and behaviors. In _____ (Chair), *Effects of a two-generation program on mothers and children: The Teenage Parent Demonstration Program.* Symposium conducted at the 2nd National Head Start Research Conference, Washington, D.C.

Abidin. R. R. (1983). *Parenting stress index-Manual.* Charlottesville, VA: Pediatric Psychology Press.

Adams, G. (1990). *Sources of support for adolescent mothers.* Washington, DC: Congressional Budget Office.

Affleck, G., Allen, D., McGrade, D. J., and McQueeney, M. (1982). Home environments of developmentally disabled infants as a function of parent and infant characteristics. *American Journal of Mental Deficiency, 86*(5), 445-452.

Anderson, B. A., Silver, B. D., and Abramson, P. R. (1988a). The effects of race of the interviewer on measures of electoral participation by blacks in SRC national election studies. *Public Opinion Quarterly, 52,* 53-83.

———. (1988b). The effects of the race of the interviewer on race-related attitudes of black respondents in SRC/CPS national election studies. *Public Opinion Quarterly, 52*, 289-324.

Andrews, F. M. (1984). Construct validity and error components of survey measures: A structural modeling approach. *Public Opinion Quarterly, 48,* 409-442.

Baker, P. C., Keck, C. K, Most, F. L., and Quinlan, S. V. (1993). *NLSY child handbook, revised edition: A guide to the 1986-1990 National Longitudinal Survey of Youth child data.* Columbus: Center for Human Resource Research, Ohio State University.

Baker, P. C., and Mott, F. L. (1989). *NLSY child handbook 1989: A guide and resource document for the National Longitudinal Study of Youth 1986 child data.* Columbus: Center for Human Resource Research, Ohio State University.

Bane, M. J., and Ellwood, D. (1983). *The dynamics of dependence and the routes to self- sufficiency.* Final report to the U.S. Department of Health and Human Services. Cambridge, MA: Harvard University, Kennedy School of Government.

Barnard, K. E. (1988). *Difficult Life Circumstances scale.* Seattle, WA: University of Washington.

Baumrind, D. (1971). Harmonious parents and their preschool children. *Developmental Psychology, 4*(1), 99-102.

Baumrind, D. (1989). Rearing competent children. In W. Damon (Ed.), *Child development today and tomorrow* (pp. 349-378). San Francisco: Jossey-Bass.

Baydar, N., and Brooks-Gunn, J. (1991). Effects of maternal employment and child care arrangement on preschoolers' cognitive and behavioral outcomes: Evidence from the children of the National Longitudinal Survey of Youth. *Developmental Psychology, 27*(6), 932-945.

Baydar, N., Brooks-Gunn, J., and Furstenberg, F. F. (1993). Early warning signs of functional illiteracy: Predictors in childhood and adolescence. *Child Development, 64*, 815-829.

Belsky, J. (1984).The determinants of parenting: A process model. *Child Development, 55*, 83-96.

448

Belsky, J., and Eggebeen, D. J. (1991). Early and extensive maternal employment and young children's socioemotional development: Children of the National Longitudinal Survey of Youth. *Journal of Marriage and the Family, 53*(4), 1083-1099.

Berlin, L. J., Brooks-Gunn, J., Spiker, D., and Zaslow, M. J. (1995). Examining observational measures of emotional support and cognitive stimulation in black and white mothers of preschoolers. *Journal of Family Issues,16*(5), 664-686.

Berlin, L. J., and Zaslow, M. J. (1996, June). Examining observational measures of emotional support and cognitive stimulation in poor, African-American adolescent mothers and their preschool-aged children: Implications for evaluating early intervention program In L. J. Berlin (Chair) *Parenting in early intervention programs: Program goals, effects, and measurement strategies.* Symposium conducted at the Head Start National Research Conference, Washington, DC.

Billiet, J., and Loosveldt, G. (1988). Improvement of the quality of responses to factual survey questions by interviewer training. *Public Opinion Quarterly, 52,* 190-211.

Blair, E. (1980). Using practice interviews to predict interviewer performance. *Public Opinion Quarterly, 44*, 257-260.

Blixt, S. (1994). Information wanted on behavior coding. *Survey Research, 25*,8.

Bornstein, M. H. (1995). *Handbook of parenting, Vols I - IV.* National Institute of Child Health and Human Development.

Bowlby, J. (1969/1982). *Attachment and loss, Vol. 1: Attachment.* New York: Basic Books.

Bowles, S., and Gintis, H. (1976). *Schooling in capitalist America.* New York: Basic Books.

Bracken, B. A. (1984). *Bracken Basic Concept Scale, examiner's manual.* The Psychological Corporation; Harcourt Brace.

Bradburn, N. M. (1985). Book review: Charles F. Turner and Elizabeth Martin, Eds., Surveying subjective phenomena. *Public Opinion Quarterly, 49,* 133-135.

Bradburn, N. M., and Sudman, S. (1979). *Improving interview method and questionnaire design.* San Francisco: Jossey-Bass.

Bradley, R. H. and Caldwell, B. M. (1979). Home observation for measurement of the environment: A revision of the pre-school scale. *American Journal of Mental Deficiency, 84*, 235-244.

———. (1980). The relation of home environment, cognitive competence and IQ among males and females. *Child Development, 51*, 1140-1148.

———. (1984). The relation of infants' home environment to achievement test performance in first grade: A follow-up study. *Child Development, 55,* 803-809.

Bradley, R. H., Caldwell, B. M., and Rock, S. L. (1988). Home environment and school performance: A ten-year follow-up and examination of three models of environmental action. *Child Development, 59*, 852-867.

Bradley, R. H., Caldwell, B. M., Rock, S. L., Ramey, C. T., Barnard, K. E., Gray, C., Hammond, M. A., Mitchell, S., Gottfried, A. W., Siegel, L., and Johnson, D. L. (1989). Home environment and cognitive development in the first three years of life: A collaborative study involving six sites and three ethnic groups in North America. *Developmental Psychology, 25*, 217-235.

Bradley, R. H., Mundfrom, D. J., Whiteside, L., Casey, P. H., and Barnett, K. (1994). A factor analytic study of the infant-toddler and early childhood versions of the HOME Inventory administered to White, Black, and Hispanic American parents of children born preterm. *Child Development, 65*, 880-888.

Bronfenbrenner, U. (1979). *The ecology of human development: Experiments by nature and design*. Cambridge, MA: Harvard University Press.

Bronfenbrenner, U. (1986). Ecology of the family as a context for human development: Research perspectives. *Developmental Psychology, 22*, 723-742.

Brooks-Gunn, J. , and Berlin, L. (1993, November). *A consideration of self-sufficiency and parenting in the context of the Teenage Parent Demonstration.* Paper presented at the Second National Head Start Research Conference.

Brooks-Gunn, J., and Furstenberg, Jr., F.F. (1986). The children of adolescent mothers: Physical, academic and psychological outcomes. *Developmental Review, 6*, 224-251.

Brown, B. B., and Stagner, M. (Eds.) (1996). *Trends in the well-being of America's children and youth: 1996.* Washington, DC: U.S. Government Printing Office.

Bryant, W. K., and Zick, C. D. (1993). *A progress report on parent-child shared time: Child training in the household.* Paper presented at the 1993 Population Association of America's Conference, Cincinnati, OH.

Bullock, M. (1995). Macro trends evident at SRCD meeting. *APA Monitor.* August 21.

Caldwell, B. M., and Bradley R. H. (1978). *Administration manual: Home observation for measurement of the environment.* Little Rock, AK: University of Arkansas.

Caldwell, B. M., and Bradley, R. H. (1984). *Home observation for measurement of the environment.* Little Rock, AK.

Cannell, C. F. (1988). Book review: Elliot G. Mishler, Research interviewing. *Public Opinion Quarterly, 52*, 265-269.

Cannell, C. F., Lawson, S., and Hausser, D. (1975). *A technique for evaluating interviewer performance.* Ann Arbor: Survey Research Center, University of Michigan.

Cannell, C. F., Miller, P. V., and Oksenberg, L. (1981). Research on interviewing techniques. In S. Leinhardt (Ed.), *Sociological methodology* (pp. 389-437). San Francisco: Jossey-Bass.

Carle, E. (1969). *The very hungry caterpillar.* New York: Philomel Books.

Chase-Lansdale, P. L., Brooks-Gunn, J., and Paikoff, R. L. (1991). Research and programs for adolescent mothers: Missing links and future promises. *Family Relations, 40*, 396-403

Chase-Lansdale, P. L., Mott, F. L., Brooks-Gunn, J., and Phillips, D. (1991). Children of the NLSY: A unique research opportunity? *Developmental Psychology, 27*, 918-931.

Chase-Lansdale, P. L., Brooks-Gunn, J., and Zamsky, E. S. (1994). Young African-American multigenerational families in poverty: Quality of mothering and grandmothering. *Child Development, 65*, 373-393.

Clark, H. H., and Schober, M. F. (1992). Asking questions and influencing answers. In J. M. Tanur (Ed.), *Questions about questions: Inquiries into the cognitive bases of surveys* (pp. 15-48). New York: Russell Sage Foundation.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Colletta, N. D. (1981). Social support and the risk of maternal rejection by adolescent mothers. *The Journal of Psychology, 109,* 191-197.

Crnic, K., Greenberg, M. T., Ragozin, A., Robinson, N., and Basham, R. (1983). Effects of stress and social support on mothers and premature and full-term infants. *Child Development, 54*, 209-217.

Crockenberg, S. (1987a). Support for adolescent mothers during the postnatal period: Theory and research. In C.F. Zachariah Boujkydis (Ed.), *Research on support for parents and infants in the postnatal period*. Norwood, NJ: Ablex.

Crockenberg, S. (1987b). Predictors and correlates of anger toward and punitive control of toddlers by adolescent mothers. *Child Development, 58*, 964-975.

D'Amico, R. J., Haurin, R. J., and Mott, F. L. (1983). The effects of mothers' employment on adolescent and early adult outcomes of young men and women. In C. D. Hayes and S. B. Kamerman (Eds.), *Children of working parents: Experiences and outcomes* (pp. 130-219). Washington, DC: National Academy Press.

DeMaio, T. J. (1991). Book review: Floyd J. Fowler, Jr., and Thomas W. Mangione. Standardized survey interviewing: Minimizing interviewer-related error. *Public Opinion Quarterly, 55,* 467-468.

Denham, S. A., Renwick, S. M., and Holt, R. W. (1991). Working and playing together: Prediction of preschool social-emotional competence from mother-child interaction. *Child Development, 62,* 242-249.

Desai, S., Chase-Lansdale, P. L., and Michael, R. T. (1989). Mother or market? Effects of maternal employment on the intellectual ability of 4-year-old children. *Demography, 26*(4), 545-561.

De Temple, J. M. (1994). *Book reading styles of low-income mothers with preschoolers and children's later literacy skills*. Unpublished doctoral dissertation, Harvard Graduate School of Education, Cambridge, MA.

De Temple, J. M., and Beals, D. E. (1991). Family talk: Sources of support for the development of decontextualized language skills. *Journal of Research in Childhood Education, 6* (1), 11-19.

Dickinson, D. K., De Temple, J. M., Hirschler, J. A., and Smith, M. W. (1992). Book reading with preschoolers: Co-construction of text at home and at school. *Early Childhood Research Quarterly, 7*(3), 323-346.

Dillman, D. (1995). Why innovation is difficult in government surveys. *Survey Research 26* (1-2), 3-6.

Dippo, C. S., and Norwood, J. L. (1992). A review of research at the Bureau of Labor Statistics. In J. M. Tanur (Ed.), *Questions about questions: Inquiries into the cognitive bases of surveys* (pp. 271-290). New York: Russell Sage Foundation.

Dishion, T. J. (1990). The family ecology of boys' peer relations in middle childhood. *Child Development, 61*, 874-892.

Downey, G., and Coyne, J. C. (1990). Children of depressed parents: An integrative review. *Psychological Bulletin, 108,* 50-76.

Duncan, G., Brooks-Gunn, J., and Klebanov, P. K. (1994). Economic deprivation and early childhood development. *Child Development, 65*, 296-318.

Egeland, B., and Brunnquell, D. (1979). An at-risk approach to the study of child abuse: Some preliminary findings. *Journal of the American Academy of Child Psychiatry, 8*, 219-235.

Egeland, B., and Kreutzer, T. (1991). A longitudinal study of the effects of maternal stress and protective factors on the development of high risk children. In E. M. Cummings, A. L. Greene, and K. H. Karraker (Eds.), *Life-span developmental psychology: Perspectives on stress and coping* (pp. 61-85). Hillsdale, NJ: Lawrence Erlbaum Associates.

Egeland, B., Weinfield, N., Hiester, M., Lawrence, C., Pierce, S., Chippendale, K., and Powell, J. (1995). *Teaching tasks administration and scoring manual.* University of Minnesota.

Egeland, B., Zaslow, M., Ogawa, J., and Weinfield, N. (1995, March). *The effects of the New Chance Program on mother-child interaction*. Paper presented at the meeting of the Society for Research in Child Development, Indianapolis.

Elardo, R. D., and Bradley, R. H. (1981). The home observation for measurement of the environment (HOME) scale: A review of research. *Developmental Review, 1*, 113-145.

Eldred, C. A. (1994). *The racial match of interviewer and respondent as a factor in self-reported respondent comfort in a non-traditional survey setting*. Paper presented at the 49th annual conference of the American Association for Public Opinion Research, Danvers, MA.

Erickson, M. F., Sroufe, L. A., and Egeland, B. (1985). The relationship between quality of attachment and behavior problems in preschool in a high-risk sample. In I. Bretherton and E. Waters (Eds.), *Growing points in attachment theory and research. Monographs of the Society for Research in Child Development, 50* (1-2), 147-166.

Farber, E., and Egeland, B. (1987). Invulnerability among abused and neglected children. In E. J. Anthony and B. Cohler (Eds.), *The invulnerable child* (pp. 253-288). New York: Guilford Press.

Fink, B. (1995). Providing quality child day care in a comprehensive program for disadvantaged young mothers and their children. *Child Welfare, LXXIV*, 1109-1134.

Fowler, F. J., Jr. (1995). *Improving survey questions: Design and evaluation.* Thousand Oaks, CA: Sage Publications.

Fowler, F. J., Jr., and Mangione, T. W. (1991). *Standardized survey interviewing: Minimizing interviewer-related error*. Newbury Park, CA: Sage Publications.

Friedman, S. L. (Chair). (1995, March). *Child care in the 1990's: The NICHD study of early child care*. Symposium at the Society for Research in Child Development, Indianapolis.

Furstenberg, F. F., Brooks-Gunn, J., and Morgan, P. S. (1987). *Adolescent mothers in later life.* Cambridge, England: Cambridge University Press.

Gallaway, C., and Richards, B. (1994). *Input and interaction in language acquisition*. Cambridge, England: Cambridge University Press.

Garcia Coll, C. T. (1990). Developmental outcome of minority infants: A process-oriented look at our beginnings. *Child Development, 61*, 270-289.

Garcia Coll, C. T., Hoffman, J., and Oh, W. (1987). The social ecology and early parenting of Caucasian adolescent mothers. *Child Development, 58*, 955-963.

Goldfield, B., and Snow, C. E. (1984). Reading books with children: The mechanics of parental influence on reading achievement. In J. Flood (Ed.), *Promoting reading comprehension*. Newark: International Reading Association.

Granger, R. C. (1994, October). *The policy implication of recent findings from the New Chance Demonstration, Ohio's Learning, Earning, and Parenting (LEAP) Program in Cleveland, and the Teenage Parent Demonstration (TPD).* Paper presented at the annual meeting of the Association for Public Policy and Management, Chicago.

Granger, R. C. (1995). Testimony of Robert C. Granger before the Senate Committee on Finance on March 14, 1995.

Groves, R. M. (1987). Research on survey data quality. *Public Opinion Quarterly, 51,* S156-S172.

Groves, R. M., and Magilavy, L. J. (1986). Measuring and explaining interviewer effects in centralized telephone surveys. *Public Opinion Quarterly, 50,* 251-266.

Harms, T., and Clifford, R. M. (1980). *The early childhood environment rating scale.* New York: Teachers College Press.

Harms, T., Cryer, D., and Clifford, R. M. (1986). *Infant-toddler environment rating scale.* New York: Teachers College Press.

Harrington, G., Block, J. H., and Block, J. (1978). Intolerance of ambiguity in preschool children: Psychometric considerations, behavioral manifestations, and parental correlates. *Developmental Psychology, 14,* 242-256.

Hart, C. H., Ladd, G. W., and Burleson, B. R. (1990). Children's expectations of the outcomes of social strategies: Relations with sociometric status and maternal disciplinary styles. *Child Development, 61,* 127-137.

Haskins, R. (1985). Public school aggression among children with varying day-care experience. *Child Development, 56,* 689-703.

Hauser, R. M., and Mossel, P. A. (1985). Fraternal resemblance in educational attainment and occupational status. *American Journal of Sociology, 91,* 650-673.

Hauser-Cram, P., Pierson, D. E., Walker, D. K., and Tivnan, T. (Eds.) (1991). *Early education in the public schools: Lessons from a comprehensive birth to kindergarten program.* San Francisco: Jossey-Bass.

Hayes, C. (Ed.). (1987). *Risking the future (Vol. I).* Washington, DC.: National Academy Press.

Heath, S. B. (1983). *Ways with words.* New York: Cambridge University Press.

Herot, C. (1996). *The wheels on the bus go round and round.* Unpublished doctoral dissertation, Harvard Graduate School of Education, Cambridge, MA.

Hetherington, E. M., Cox, M., and Cox, R. (1985). Long-term effects of divorce and remarriage on the adjustment of children. *Journal of the American Academy of Child Psychiatry, 24,* 518-530.

Hill, C. R., and Stafford, F. P. (1974). Allocation of time to preschool children and educational opportunity. *Journal of Human Resources, 9,* 323-343.

Holden, G. W., and Edwards, L. A. (1989). Parental attitudes toward child rearing: Instruments, issues, and implications. *Psychological Bulletin, 106,* 29-58.

Hollenbeck, A. R. (1978). Early infant home environments: Validation of the HOME observation for measurement of the environmental inventory. *Developmental Psychology, 14*(4), 416- 418.

Klebanov, P. K., Brooks-Gunn, J., and Duncan, G. J. (1994). Does neighborhood and family poverty affect mothers' parenting, mental health, and social support? *Journal of Marriage and the Family, 56,* 441-455.

Kohn, M. L., and Schooler, C. (1982). Job conditions and personality: A longitudinal assessment of their reciprocal effects. *American Journal of Sociology, 87*, 1257–1286.

________. (1983). *Work and personality: An inquiry into the impact of social stratification.* Norwood, NJ: Ablex.

Kurtz, B. E., Borkowski, J. G., and Deshmukh, K. (1986). Metamemory and learning in Maharashtrian children: Influences from home and school. (Rev. ed.) *Journal of Genetic Psychology, 149*(3), 363-376.

Landis, J. R., and Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*, 159-174.

Landry, S. H., Chapieski, M. L., Richardson, M. A., Palmer, J., and Hall, S. (1990). The social competence of children born prematurely: Effects of medical complications and parent behaviors. *Child Development, 61*, 1605-1616.

Maccoby, E. M., and Martin, J. M. (1983). Socialization in the context of the family: Parent-child interaction. In P. H. Mussen (Ed.), *Handbook of child psychology, Vol. IV* (pp. 1-101). New York: Wiley.

MacWhinney, B., and Snow, C. E. (1985). The child language data exchange system. *Journal of Child Language, 12*, 271-295.

Mariner, C., and Zaslow, M. (1997). *Don't leave HOME without this: Working paper on the use of the HOME-SF as a survey measure of parenting*. Washington, D. C.: Child Trends, Inc.

McLoyd, V. C. (1990). The impact of economic hardship on Black families and children: Psychological distress, parenting, and socioemotional development. *Child Development, 61*, 311-346.

Menaghan, G., and Parcel, T. L. (1995). Social sources of change in children's home environments: The effects of parental occupational experiences and family conditions. *Journal of Marriage and the Family, 57*, 69-84.

Miller, P. V., and Groves, R. M. (1985). Matching survey responses to official records: An exploration of validity in victimization reporting. *Public Opinion Quarterly, 4*, 366-380.

Mishler, E. G. (1986). *Research interviewing*. Cambridge, MA: Harvard University Press.

Moore, K. A. (1986, May). *Children of teen parents: Heterogeneity of outcomes*. Final report to the Center for Population Research, NICHD, Grant No. HD-184272.

Moore, K. A. (1990). *Facts at a glance — Annual newsletter on teen pregnancy*. Washington, DC: Child Trends, Inc.

Moore, K. A, and Snyder, N. O. (1991). Cognitive development among firstborn children of adolescent mothers. *American Sociological Review, 56*(5), 612-624.

Moore, K. A., Zaslow, M. J., Coiro, M. J., Miller, S. M., and Magenheim, E. B. (1995). *The JOBS Evaluation: How Well Are They Faring? AFDC Families with Preschool-Aged Children in Atlanta at the Outset of the JOBS Evaluation*. U.S Department of Health and Human Services and U.S. Department of Education. Washington, DC: U.S. Department of Health and Human Services, Office of the Assistant Secretary for Planning and Evaluation.

Oksenberg, L., Coleman, L., and Cannell, C. F. (1986). Interviewers' voices and refusal rates in telephone surveys. *Public Opinion Quarterly, 50*, 97-111.

O'Rourke, A. B. (1990). *Reading aloud to children preschool to grade 6*. Unpublished doctoral dissertation, Harvard Graduate School of Education, Cambridge, MA.

Parks, P. L., and Bradley, R. H. (1991). The interaction of home environment as a mediating construct between SES and child outcomes. Columbus, OH: Ohio State University, Department of Sociology.

Patterson, C. J., Cohn, D. A., and Kao, B. T. (1989). Maternal warmth as a protective factor against risks associated with peer rejection among children. *Developmental Psychopathology, 1*, 21-38.

Pearlin, L. I., Menaghan, E. G., Lieberman, M. A., and Mullan, J. T. (1981). The stress process. *Journal of Health and Social Behavior, 22*, 337-356.

Pianta, R., and Egeland, B. (1994). Predictors of instability in children's mental test performance at 24, 48 and 96 months. *Intelligence, 18*(2), 145-163.

Pianta, R., Erickson, M. F., Wagner, N., Kreutzer, T., and Egeland, B. (1990). Early predictors of referral for special services: Child-based measures versus mother-child interaction. *School Psychology Review, 19*(2), 240-250.

Polit, D. F. (1996a). *Parenting and child outcome measures in the New Chance 42-month survey*. New York: Manpower Demonstration Research Corporation.

Polit, D. F. (1996b). *Self-administered teacher questionnaire in the New Chance 42-month survey*. New York: Manpower Demonstration Research Corporation.

Presser, S., and Zhao, S. (1992). Attributes of questions and interviewers as correlates of interviewing performance. *Public Opinion Quarterly 56*, 236-240.

Quint, J. C., Bos, J. M., and Polit, D. F. (1997). *New Chance: Final report on a comprehensive program for young mothers in poverty and their children*. New York: Manpower Demonstration Research Corporation.

Quint, J. C., and Egeland, B. (1995) Two-generation programs for families in poverty: A new intervention strategy. In S. Smith (Ed.), *Two-generation programs for families in poverty*. Norwood, NJ: Ablex.

Quint, J. C., Fink, B. L., and Rowser, S. L. (1991). *New Chance: Implementing a comprehensive program for disadvantaged young mothers and their children*. New York: Manpower Demonstration Research Corporation.

Quint, J. C., and Musick, J. S. (1994). *Lives of promise, lives of pain: Young mothers after New Chance*. New York: Manpower Demonstration Research Corporation.

Quint, J. C., Polit, D. F., Bos, H., and Cave, G. (1994). *New Chance: Interim findings on a comprehensive program for disadvantaged young mothers and their children*. New York: Manpower Demonstration Research Corporation.

Radloff, L. S. (1977). The CES-D scale: A self-report depression scale for research in the general population. *Applied Psychological Measurement, 1*, 385-401.

Rahe, D. F. (1984). *Interaction patterns between children and mothers on teaching tasks at age 42 months: Antecedents in attachment history, intellectual correlates, and consequences on children's socio-emotional functioning*. Unpublished doctoral dissertation, University of Minnesota.

Ramey, C. T. (1992). High risk children and IQ: Altering intergenerational patterns. *Intelligence, 15*, 239-256.

Ramey, C. T., Bryant, D. M., Wasik, B. H., Sparling, J. J., Fendt, K. H., and LaVange, L. M. (1992). The infant health and development program for low birthweight, premature infants: Program elements, family participation, and child intelligence. *Pediatrics, 89*, 454-465.

Ramey, C. T., Ramey, S. L., Gaines, K. R., and Blair, C. (1995). Two-generation early intervention programs: A child development perspective. In S. Smith and I. E. Sigel (Eds.), *Two-generation programs for families in poverty: A new intervention strategy* (pp. 199–228). Norwood, NJ: Ablex.

Ramey, C. T., Yates, and Short (1984). The plasticity of intellectual development: Insights from preventive intervention. *Child Development, 55*, 1913-1925.

Reese, S. D., Danielson, W. A, Shoemaker, P .J., Chang, T., Hsu, H. (1986). Ethnicity-of-interviewer effects among Mexican-Americans and Anglos. *Public Opinion Quarterly, 50*, 563-572.

Reiss, D., Plomin, R., Hetherington, E. M., Howe, G. W., Rovine, M., Tryon, A., and Hagan, S. M. (1994). The separate worlds of teenage siblings: An introduction to the study of the nonshared environment and adolescent development. E. M. Hetherington, D. Reiss, and R. Plomin (Eds.), *Separate social worlds of siblings*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Renken, B., Egeland, B., Marvinney, D., Mangelsdorf, S. and Sroufe, L. A. (1989). Early childhood antecedents of aggression and passive-withdrawal in early elementary school. *Journal of Personality, 57*, 257-281.

Richman, N., Stevenson, J., and Graham, P. J. (1982). *Preschool to school: A behavioral study*. London: Academic.

Rotter, J. B. (1966). Generalized expectancies for internal versus external control of reinforcement. *Psychological Monographs, 66*(80).

Rutter, M. (1987). Psychological resilience and protective mechanisms. *American Journal of Orthopsychiatry, 57*, 316-331.

Savin-Williams, R. C. and Berndt, T. J. (1990). Friendship and peer relations. In S. Feldman and G. R. Elliott (Eds.), *At the threshold: The developing adolescent* (pp. 277-307). Cambridge, MA: Harvard University Press.

Schilmoeller, G. L., and Baranowski, M. D. (1985). Childrearing of firstborns by adolescent and older mothers. *Adolescence, 10* (80), 805-822.

Sewell, W. H., and Hauser, R. M. (1976). Causes and consequences of higher education: Models of the status attainment process. In W. H. Sewell, R. M. Hauser, and D. L. Featherman (Eds.), *Schooling and achievement in American society* (pp. 9-27). New York: Academic Press.

Singer, E., Frankel, M. R., and Glassman, M. B. (1983). The effect of interviewer characteristics and expectations on response. *Public Opinion Quarterly, 47*, 68-83.

Smith, S. (1993). Discussant's comments, symposium entitled *Evaluating two-generation interventions: Recommendations for design, analysis, and field implementation* . At Second National Head Start Research Conference, Washington, DC, November, 1993.

Smith, S. (Ed.) (1995). *Two-generation programs for families in poverty: A new intervention strategy*. Norwood, NJ: Ablex.

Smith, S., and Zaslow, M. (1995). Rationale and policy context for two-generation interventions. In S. Smith (Ed.), *Two-generation programs for families in poverty: A new intervention strategy* (pp. 1–36). Norwood, NJ: Ablex.

Snow, C. E., Barnes, W. S., Chandler, J., Goodman, I. F., and Hemphill, J. (1991*). Unfulfilled expectations: Home and school influences on literacy*. Cambridge, MA: Harvard University Press.

Snow, C. E., Cancino, H., Gonzalez, P.,and Shriberg, E. (1989). Giving formal definitions: An oral correlate of literacy. In D. Bloome (Ed.), *Classrooms and literacy*. Norwood, NJ: Ablex.

Snow, C. E., Tabors, P. O., Nicholson, P., and Kurland, B. F. (1995). SHELL: Oral language and early literacy skills in kindergarten and first grade children. *Journal of Research in Childhood Education, 10*(1), 37-48.

Sroufe, L. A. (1979). The coherence of individual development. *American Psychologist, 34*, 834- 841.

Sroufe, L. A. (1988). The role of infant-caregiver attachment in development. In J. Belsky and T. Nezworski (Eds.), *Clinical Implications of Attachment* (pp. 18-38). Hillsdale, NJ: Lawrence Erlbaum Associates.

Sroufe, L. A., Egeland, B., and Kreutzer, T. (1990). The fate of early experience following developmental change: Longitudinal approaches to individual adaptation in childhood. *Child Development, 61*, 1363-1373.

Steinberg, L. (1990). Autonomy, conflict, and harmony in the family relationship. In S. Feldman and G. R. Elliott (Eds.), *At the threshold: The developing adolescent* (pp. 255-276). Cambridge, MA: Harvard University Press.

Stevens, J. H. (1988). Social support, locus of control, and parenting in three low-income groups of mothers: Black teenagers, black adults, and white adults. *Child Development, 59*, 635-642.

Sugland, B. W., Zaslow, M. J., Smith, J. R., Brooks-Gunn, J., Coates, D., Blumenthal, C., Moore, K. A., Griffin, T., and Bradley, R. (1995). The early childhood HOME inventory and HOME-short form in differing racial/ethnic groups: Are there differences in underlying structure, internal consistency of subscales, and patterns of prediction? *Journal of Family Issues, 16*(5), 632-663.

Tanur, J. M. (Ed.) (1992). *Questions about questions: Inquiries into the cognitive bases of surveys*. New York: Russell Sage Foundation.

Tapia-Uribe, F. M. (1988). *Women's schooling, fertility and child survival in a Mexican village*. Unpublished doctoral dissertation, Harvard Graduate School of Education, Cambridge, MA.

Teale, W. H., and Martinez, M. (1986). Teachers' storybook reading styles: Evidence and implications. *Reading Education in Texas, 2*, 7-16.

Tucker, C. (1983). Interviewer effects in telephone surveys. *Public Opinion Quarterly, 47*, 84-95.

Turner, C. F., Lessler, J. T., and Gfroerer, J. C. (Eds.) (1992). *Survey measurement of drug use: Methodological studies*. Washington, DC: U.S. Government Printing Office.

Unger, D. G., and Wandersman, L. P. (1988). The relation of family and partner support to the adjustment of adolescent mothers. *Child Development, 59*, 1056-1060.

Walker, T. B., Rodriguez, G. G., Johnson, D. L., and Cortez, C. C. (1995). Avance parent-child education program. In S. Smith (Ed.), *Two-generation programs for families in poverty: A new intervention strategy* (pp. 67-90). Norwood, NJ: Ablex.

Watson, R. (1989). Literate discourse and cognitive organization: Some relations between parents' talk and 3-year-olds' thought. *Applied Psycholinguistics, 10*, 221-236.

Wells, G. (1985). Preschool literacy-related activities and success in school. In D. R. Olson, N. Torrance, and A. Hildyard (Eds.), *Literacy, language and learning: The nature and consequences of reading and writing* (pp. 229-255). New York: Cambridge University Press.

Whitehurst, G. J., Galco, F. L., Lonigan C. J., and Fischel, J. E. et al. (1988). Accelerating language development through picture book reading. *Developmental Psychology, 24*(4), 552-559.

Zaslow, M. J., Oldham, E., Moore, K. A., and Magenheim, E. (1997, April). *The implications of participation in community-based child care for the cognitive and social development of children from welfare families*. Paper presented at the meetings of the Society for Research in Child Development, Washington, DC.

Zill, N. (1985). *Behavior Problem Scales developed from the 1981 Child Health supplement to the National Health Interview Survey*. Washington, DC: Child Trends.

Selected Publications on MDRC Projects

Programs for Teenage Parents

The New Chance Demonstration

A test of a comprehensive program of services that seeks to improve the economic status and general well-being of a group of highly disadvantaged young women and their children.

New Chance: Implementing a Comprehensive Program for Disadvantaged Young Mothers and Their Children. 1991. Janet Quint, Barbara Fink, Sharon Rowser.

Lives of Promise, Lives of Pain: Young Mothers After New Chance. Monograph. 1994. Janet Quint, Judith Musick, with Joyce Ladner.

New Chance: Interim Findings on a Comprehensive Program for Disadvantaged Young Mothers and Their Children. 1994. Janet Quint, Denise Polit, Hans Bos, George Cave.

New Chance: The Cost Analysis of a Comprehensive Program for Disadvantaged Young Mothers and Their Children. Working Paper. 1994. Barbara Fink.

New Chance: Final Report on a Comprehensive Program for Young Mothers in Poverty and Their Children. 1997. Janet Quint, Johannes M. Bos, Denise Polit.

Parenting Behavior in a Sample of Young Mothers in Poverty: Results of the New Chance Observational Study. 1998. Martha Zaslow, Carolyn Eldred, editors.

The LEAP Evaluation

An evaluation of Ohio's Learning, Earning, and Parenting (LEAP) Program, which uses financial incentives to encourage teenage parents on welfare to stay in or return to school.

LEAP: Implementing a Welfare Initiative to Improve School Attendance Among Teenage Parents. 1991. Dan Bloom, Hilary Kopp, David Long, Denise Polit.

LEAP: Interim Findings on a Welfare Initiative to Improve School Attendance Among Teenage Parents. 1993. Dan Bloom, Veronica Fellerath, David Long, Robert Wood.

LEAP: The Educational Effects of LEAP and Enhanced Services in Cleveland. 1994. David Long, Robert Wood, Hilary Kopp.

LEAP: Three-Year Impacts of Ohio's Welfare Initiative to Improve School Attendance Among Teenage Parents. 1996. David Long, Judith Gueron, Robert Wood, Rebecca Fisher, Veronica Fellerath.

LEAP: Final Report on Ohio's Welfare Initiative to Improve School Attendance Among Teenage Parents. 1997. Johannes Bos, Veronica Fellerath.

Project Redirection

A test of a comprehensive program of services for pregnant and parenting teenagers.

The Challenge of Serving Teenage Mothers: Lessons from Project Redirection. Monograph. 1988. Denise Polit, Janet Quint, James Riccio.

The Community Service Projects

A test of a New York State teenage pregnancy prevention and services initiative.

The Community Service Projects: Final Report on a New York State Adolescent Pregnancy Prevention and Services Program. 1988. Cynthia Guy, Lawrence Bailis, David Palasits, Kay Sherwood.

Note: For works not published by MDRC, the publisher's name is shown in parentheses.

Reforming Welfare

Books and Monographs

Reforming Welfare with Work (Ford Foundation). Monograph. 1987. Judith Gueron. A review of welfare-to-work initiatives in five states.

From Welfare to Work (Russell Sage Foundation). Book. 1991. Judith Gueron, Edward Pauly. A synthesis of research findings on the effectiveness of welfare-to-work programs. Chapter 1, which is the summary of the book, is also published separately by MDRC.

Five Years After: The Long-Term Effects of Welfare-to-Work Programs (Russell Sage Foundation). Book. 1995. Daniel Friedlander, Gary Burtless. An analysis of five-year follow-up data on four welfare-to-work programs.

After AFDC: Welfare-to-Work Choices and Challenges for States. Book. 1997. Dan Bloom. A summary and synthesis of lessons derived from studies of welfare reform programs.

ReWORKing Welfare: Technical Assistance for States and Localities

After AFDC: Welfare-to-Work Choices and Challenges for States. See under Books and Monographs.

Changing to a Work First Strategy: Lessons from Los Angeles County's GAIN Program for Welfare Recipients. 1997. Evan Weissman.

Work First: How to Implement an Employment-Focused Approach to Welfare Reform. 1997. Amy Brown.

Connections to Work Project

Alternative approaches to help welfare recipients and other low-income populations access and secure jobs.

Tulsa's IndEx Program: A Business-Led Initiative for Welfare Reform and Economic Development. 1997. Maria Buck.

Working Papers

Working Papers related to a specific project are listed under that project.

Learning from the Voices of Mothers: Single Mothers' Perceptions of the Trade-offs Between Welfare and Work. 1993. LaDonna Pavetti.

Unpaid Work Experience for Welfare Recipients: Findings and Lessons from MDRC Research. 1993. Thomas Brock, David Butler, David Long.

From Welfare to Work Among Lone Parents in Britain: Lessons for America. 1996. James Riccio.

Papers for Practitioners

Assessing JOBS Participants: Issues and Trade-offs. 1992. Patricia Auspos, Kay Sherwood.

Linking Welfare and Education: A Study of New Programs in Five States. 1992. Edward Pauly, David Long, Karin Martinson.

Improving the Productivity of JOBS Programs. 1993. Eugene Bardach.

Reports and Other Publications

Time-Limited Welfare

Florida's Family Transition Program

A study of Florida's time-limited welfare program.

The Family Transition Program: An Early Implementation Report on Florida's Time-Limited Welfare Initiative. 1995. Dan Bloom.

BEST COPY AVAILABLE

The Family Transition Program: Implementation and Early Impacts of Florida's Initial Time-Limited Welfare Program. 1997. Dan Bloom, James Kemple, Robin Rogers-Dillon.

The Cross-State Study of Time-Limited Welfare

An examination of the implementation of some of the first state-initiated time-limited welfare programs.

Implementing Time-Limited Welfare: Early Experiences in Three States. 1995. Dan Bloom, David Butler.

The View from the Field: As Time Limits Approach, Welfare Recipients and Staff Talk About Their Attitudes and Expectations. 1997. Amy Brown, Dan Bloom. David Butler.

Making Work Pay

The Minnesota Family Investment Program (MFIP)

An evaluation of Minnesota's welfare reform initiative.

MFIP: An Early Report on Minnesota's Approach to Welfare Reform. 1995. Virginia Knox, Amy Brown, Winston Lin.

Making Welfare Work and Work Pay: Implementation and 18-Month Impacts of the Minnesota Family Investment Program. 1997. Cynthia Miller, Virginia Knox, Patricia Auspos, Jo Anna Hunter-Manns, Alan Orenstein.

The New Hope Project

A test of a neighborhood-based antipoverty program and welfare alternative operating in Milwaukee.

The New Hope Offer: Participants in the New Hope Demonstration Discuss Work, Family, and Self-Sufficiency. Working Paper. 1996. Dudley Benoit.

Who Got New Hope? Working Paper. 1997. Michael Wiseman.

Creating New Hope: Implementation of a Program to Reduce Poverty and Reform Welfare. Thomas Brock, Fred Doolittle, Veronica Fellerath, Michael Wiseman.

Canada's Self-Sufficiency Project (SSP)

A test of the effectiveness of a temporary earnings supplement on the employment and welfare receipt of public assistance recipients. Reports on the Self-Sufficiency Project are available from: Social Research and Demonstration Corporation (SRDC), 275 Slater St., Suite 900, Ottawa, Ontario K1P 5H9, Canada. Tel.: 613-237-4311; Fax: 613-237-5045. In the United States, the reports are also available from MDRC.

Making Work Pay Better Than Welfare: An Early Look at the Self-Sufficiency Project (Social Research and Demonstration Corporation). 1994. Susanna Lui-Gurr, Sheila Currie Vernon, Tod Mijanovich.

Creating an Alternative to Welfare: First-Year Findings on the Implementation, Welfare Impacts, and Costs of the Self-Sufficiency Project (Social Research and Demonstration Corporation). 1995. Tod Mijanovich, David Long.

The Struggle for Self-Sufficiency: SSP Participants Talk About Work, Welfare, and Their Futures (Social Research and Demonstration Corporation). 1995. Wendy Bancroft, Sheila Currie Vernon.

Do Financial Incentives Encourage Welfare Recipients to Work? Initial 18-Month Findings from the Self-Sufficiency Project (Social Research and Demonstration Corporation). 1996. David Card, Philip Robins.

When Work Pays Better Than Welfare: A Summary of the Self-Sufficiency Project's Implementation, Focus Group, and Initial 18-Month Impact Reports (Social Research and Demonstration Corporation). 1996.

How Important Are "Entry Effects" in Financial Incentive Programs for Welfare Recipients? Experimental Evidence from the Self-Sufficiency Project (Social Research and Demonstration Corporation). 1997. David Card, Philip Robins, Winston Lin.

JOBS Programs

The National Evaluation of Welfare-to-Work Strategies

An evaluation of welfare-to-work programs launched under the Job Opportunities and Basic Skills Training (JOBS) provisions of the Family Support Act of 1988.

From Welfare to Work (Russell Sage Foundation). See under Books and Monographs.

The Saturation Work Initiative Model in San Diego: A Five-Year Follow-up Study. 1993. Daniel Friedlander, Gayle Hamilton.

The JOBS Evaluation: Early Lessons from Seven Sites (U.S. Department of Health and Human Services [HHS]). 1994. Gayle Hamilton, Thomas Brock.

Five Years After: The Long-Term Effects of Welfare-to-Work Programs. See under Books and Monographs.

Adult Education for People on AFDC—A Synthesis of Research. (U.S. Department of Health and Human Services, Office of the Assistant Secretary for Planning and Evaluation [HHS, ASPE]). 1995. Edward Pauly.

Early Findings on Program Impacts in Three Sites (HHS, ASPE). 1995. Stephen Freedman, Daniel Friedlander.

How Well Are They Faring? AFDC Families with Preschool-Aged Children in Atlanta at the Outset of the JOBS Evaluation (HHS, ASPE). 1995. Child Trends, Inc.: Kristin Moore, Martha Zaslow, Mary Jo Coiro, Suzanne Miller, Ellen Magenheim.

Monthly Participation Rates in Three Sites and Factors Affecting Participation Levels in Welfare-to-Work Programs (HHS, ASPE). 1995. Gayle Hamilton.

Evaluating Two Welfare-to-Work Program Approaches: Two-Year Findings on the Labor Force Attachment and Human Capital Development Programs in Three Sites (HHS, Administration for Children and Families and ASPE). 1997. Gayle Hamilton, Thomas Brock, Mary Farrell, Daniel Friedlander, Kristen Harknett.

Changing to a Work First Strategy: Lessons from Los Angeles County's GAIN Program for Welfare Recipients. 1997. Evan Weissman.

Work First: How to Implement an Employment-Focused Approach to Welfare Reform. 1997. Amy Brown.

The GAIN Evaluation

An evaluation of California's Greater Avenues for Independence (GAIN) Program, the state's JOBS program.

GAIN: Planning and Early Implementation. 1987. John Wallace, David Long.

GAIN: Child Care in a Welfare Employment Initiative. 1989. Karin Martinson, James Riccio.

GAIN: Early Implementation Experiences and Lessons. 1989. James Riccio, Barbara Goldman, Gayle Hamilton, Karin Martinson, Alan Orenstein.

GAIN: Participation Patterns in Four Counties. 1991. Stephen Freedman, James Riccio.

GAIN: Program Strategies, Participation Patterns, and First-Year Impacts in Six Counties. 1992. James Riccio, Daniel Friedlander.

GAIN: Two-Year Impacts in Six Counties. 1993. Daniel Friedlander, James Riccio, Stephen Freedman.

GAIN: Basic Education in a Welfare-to-Work Program. 1994. Karin Martinson, Daniel Friedlander.

GAIN: Benefits, Costs, and Three-Year Impacts of a Welfare-to-Work Program. 1994. James Riccio, Daniel Friedlander, Stephen Freedman.

Related Studies:

The Impacts of California's GAIN Program on Different Ethnic Groups: Two-Year Findings on Earnings and AFDC Payments. Working Paper. 1994. Daniel Friedlander.

Can They All Work? A Study of the Employment Potential of Welfare Recipients in a Welfare-to-Work Program. Working Paper. 1995. James Riccio, Stephen Freedman.

Changing to a Work First Strategy: Lessons from Los Angeles County's GAIN Program for Welfare Recipients. 1997. Evan Weissman.

The Evaluation of Florida's Project Independence

An evaluation of Florida's initial JOBS program.

Florida's Project Independence: Program Implementation, Participation Patterns, and First-Year Impacts. 1994. James Kemple, Joshua Haimson.

Florida's Project Independence: Benefits, Costs, and Two-Year Impacts of Florida's JOBS Program. 1995. James Kemple, Daniel Friedlander, Veronica Fellerath.

The Family Transition Program: Implementation and Early Impacts of Florida's Initial Time-Limited Welfare Program. 1997. Dan Bloom, James Kemple, Robin Rogers-Dillon.

Other Welfare Studies

The Saturation Work Initiative Model (SWIM)

A test of the feasibility and effectiveness of an ongoing participation requirement in a welfare-to-work program.

Interim Report on the Saturation Work Initiative Model in San Diego. 1988. Gayle Hamilton.

Final Report on the Saturation Work Initiative Model in San Diego. 1989. Gayle Hamilton, Daniel Friedlander.

The Saturation Work Initiative Model in San Diego: A Five-Year Follow-up Study. 1993. Daniel Friedlander, Gayle Hamilton.

The Demonstration of State Work/Welfare Initiatives

A test of the feasibility and effectiveness of various state employment initiatives for welfare recipients.

Arizona: *Preliminary Management Lessons from the WIN Demonstration Program*. 1984. Kay Sherwood.

Arkansas: *Final Report on the WORK Program in Two Counties*. 1985. Daniel Friedlander, Gregory Hoerz, Janet Quint, James Riccio.

California: *Final Report on the San Diego Job Search and Work Experience Demonstration*. 1986. Barbara Goldman, Daniel Friedlander, David Long.

Illinois: *Final Report on Job Search and Work Experience in Cook County*. 1987. Daniel Friedlander, Stephen Freedman, Gayle Hamilton, Janet Quint.

Maine: *Final Report on the Training Opportunities in the Private Sector Program*. 1988. Patricia Auspos, George Cave, David Long.

Maryland: *Final Report on the Employment Initiatives Evaluation*. 1985. Daniel Friedlander, Gregory Hoerz, David Long, Janet Quint.

Supplemental Report on the Baltimore Options Program. 1987. Daniel Friedlander.

New Jersey: *Final Report on the Grant Diversion Project*. 1988. Stephen Freedman, Jan Bryant, George Cave.

Virginia: *Final Report on the Virginia Employment Services Program*. 1986. James Riccio, George Cave, Stephen Freedman, Marilyn Price.

West Virginia: *Final Report on the Community Work Experience Demonstrations*. 1986. Daniel Friedlander, Marjorie Erickson, Gayle Hamilton, Virginia Knox.

Other Reports on the Demonstration of State Work/Welfare Initiatives

Relationship Between Earnings and Welfare Benefits for Working Recipients: Four Area Case Studies. 1985. Barbara Goldman, Edward Cavin, Marjorie Erickson, Gayle Hamilton, Darlene Hasselbring, Sandra Reynolds.

Welfare Grant Diversion: Early Observations from Programs in Six States. 1985. Michael Bangser, James Healy, Robert Ivry.

A Survey of Participants and Worksite Supervisors in the New York City Work Experience Program. 1986. Gregory Hoerz, Karla Hanson.

Welfare Grant Diversion: Lessons and Prospects. 1986. Michael Bangser, James Healy, Robert Ivry.

Work Initiatives for Welfare Recipients: Lessons from a Multi-State Experiment. 1986. Judith Gueron.

The Subgroup/Performance Indicator Study

A study of the impacts of selected welfare-to-work programs on subgroups of the AFDC caseload.

A Study of Performance Measures and Subgroup Impacts in Three Welfare Employment Programs. 1987. Daniel Friedlander, David Long.

Subgroup Impacts and Performance Indicators for Selected Welfare Employment Programs. 1988. Daniel Friedlander.

The Self-Employment Investment Demonstration (SEID)

A test of the feasibility of operating a program to encourage self-employment among recipients of AFDC.

Self-Employment for Welfare Recipients: Implementation of the SEID Program. 1991. Cynthia Guy, Fred Doolittle, Barbara Fink.

The WIN Research Laboratory Project

A test of innovative service delivery approaches in four Work Incentive Program (WIN) offices.

Impacts of the Immediate Job Search Assistance Experiment: Louisville WIN Research Laboratory Project. 1981. Barbara Goldman.

Welfare Women in a Group Job Search Program: Their Experiences in the Louisville WIN Research Laboratory Project. 1982. Joanna Gould-Stuart.

Job Search Strategies: Lessons from the Louisville WIN Laboratory. 1983. Carl Wolfhagen, Barbara Goldman.

The Parents' Fair Share Demonstration

A demonstration aimed at reducing child poverty by increasing the job-holding, earnings, and child support payments of unemployed, noncustodial parents (usually fathers) of children receiving public assistance.

Caring and Paying: What Fathers and Mothers Say About Child Support. 1992. Frank Furstenberg, Jr., Kay Sherwood, Mercer Sullivan.

Child Support Enforcement: A Case Study. Working Paper. 1993. Dan Bloom.

Matching Opportunities to Obligations: Lessons for Child Support Reform from the Parents' Fair Share Pilot Phase. 1994. Dan Bloom, Kay Sherwood.

Low-Income Parents and the Parents' Fair Share Demonstration: An Early Qualitative Look at Low-Income Noncustodial Parents (NCPs) and How One Policy Initiative Has Attempted to Improve Their Ability to Pay Child Support. 1996. Earl Johnson, Fred Doolittle.

The National Supported Work Demonstration

A test of a transitional work experience program for four disadvantaged groups.

Summary and Findings of the National Supported Work Demonstration. 1980. MDRC Board of Directors.

The Section 3 Study

Lessons from the Field on the Implementation of Section 3 (U.S. Department of Housing and Urban Development, Office of Policy Development and Research). 1996. Maxine Bailey, Suzanne Lynn.

About MDRC

The Manpower Demonstration Research Corporation (MDRC) is a nonprofit social policy research organization founded in 1974 and located in New York City and San Francisco. Its mission is to design and rigorously field-test promising education and employment-related programs aimed at improving the well-being of disadvantaged adults and youth, and to provide policymakers and practitioners with reliable evidence on the effectiveness of social programs. Through this work, and its technical assistance to program administrators, MDRC seeks to enhance the quality of public policies and programs. MDRC actively disseminates the results of its research through its publications and through interchanges with a broad audience of policymakers and practitioners; state, local, and federal officials; program planners and operators; the funding community; educators; scholars; community and national organizations; the media; and the general public.

Over the past two decades — working in partnership with more than forty states, the federal government, scores of communities, and numerous private philanthropies — MDRC has developed and studied more than three dozen promising social policy initiatives.

MDRC

16 East 34 Street
New York, NY 10016
(212) 532-3200
www.mdrc.org

88 Kearny Street, Suite 1800
San Francisco, CA 94108
(415) 781-3800

465

BEST COPY AVAILABL